CONSTRUCTION OF A MINIMAL TILING PATH ACROSS THE EUCHROMATIC ARMS OF SORGHUM CHROMOSOME 3 AND COMPARATIVE ANALYSIS WITH THE RICE CHROMOSOME 1 PSEUDOMOLECULE

A Dissertation

by

BIN ZHOU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Molecular and Environmental Plant Sciences

CONSTRUCTION OF A MINIMAL TILING PATH ACROSS THE EUCHROMATIC ARMS OF SORGHUM CHROMOSOME 3 AND COMPARATIVE ANALYSIS WITH THE RICE CHROMOSOME 1 PSEUDOMOLECULE

A Dissertation

by

BIN ZHOU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Patricia E. Klein
Committee Members,	John E. Mullet
	Marla L. Binzel
	Carol A. Loopstra
Chair of Interdisciplinary Faculty,	Marla L. Binzel

December 2006

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Construction of a Minimal Tiling Path Across the Euchromatic Arms of Sorghum Chromosome 3 and Comparative Analysis with the Rice Chromosome 1 Pseudomolecule. (December 2006)

Bin Zhou, B.S., Huazhong Agricultural University, Hubei, China;

M.S., Huazhong Agricultural University, Hubei, China

Chair of Advisory Committee: Dr. Patricia E. Klein

Using rice chromosome 1 pseudomolecule as a reference, a minimal tiling path for the euchromatic arms of sorghum chromosome 3 was constructed, in which 23 contigs contain an estimated 57.56 Mb of DNA. A total of 409 EST-STS markers and 255 genetic markers have been mapped onto the euchromatic arms providing excellent integration of the genetic and physical maps. A total of 21 contigs containing 9 EST-STS and 35 genetic markers have been constructed across the heterochromatin block of sorghum chromosome 3 which comprise 16.57 Mb of DNA.

Macrocolinearity between sorghum chromosome 3 and rice chromosome 1 was examined based on the mapped EST-STS markers. Approximately 85% of the EST-STS markers were colinear between these two homeologous chromosomes. Estimates of recombination were also determined, which indicates the existence of recombination cold and hot spots.

Microcolinearity between sorghum chromosome 3 and rice chromosome 1 was examined at two different levels. In one case, overlapping sorghum BAC pools orthologous to a 5.1 Mb region of rice chromosome 1 were constructed and sequence skimmed. Alignment of the sorghum skim sequences to the TIGR rice gene models revealed ~62% colinearity between the two orthologous regions. In addition, colinearity between sorghum chromosome 3 and rice chromosome 5 was detected within this region which is likely due to the segmental homology between rice chromosome 1 and rice chromosome 5. Microcolinearity between sorghum and rice was also examined by comparing 2 fully sequenced sorghum chromosome 3 BAC clones to the orthologous region of rice chromosome 1. In this analysis, ~65% colinearity was detected for sorghum BAC 82G24 and ~59% colinearity was detected for sorghum BAC 181g10. Microcolinearity was largely confined to gene coding regions and sequences of exons displayed the highest percent identities. Small-scale gene rearrangements were also detected.

Finally, RT-PCR analysis was carried out between a set of colinear and non-colinear genes from sorghum and rice to determine whether the loss of colinearity between orthologous genes resulted in a change in transcriptional regulation. No direct link between loss of colinearity and expression pattern was detected in these experiments.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee, John Mullet, Marla Binzel, and Carol Loopstra; they each supported me in different ways. I am deeply grateful to my advisor, Patricia Klein, for her guidance and financial support as I pursued my degree. She also provided the encouragement, constructive comments and suggestions to help make my experience a very productive and fulfilling one. Members of the large sorghum research community at Texas A&M were helpful in various ways during the work on this project, and during my graduate studies in general. This community included: Bob Klein, Bill Rooney, Monica Menz, Christina Buchanan, Jeong-Soon Kim, Arun Sharma, Natalie Unruh, Julie McCollum, Lucy Haizlip, and fellow graduate students, Karen Harris, Jeffrey Brady and Sanghyun Lim. Above all, I thank my wife, Yingyun Wu, and my parents for their love and support through the duration of my work.

TABLE OF CONTENTS

CHAPTER Page

Page

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ABA	Abscisic Acid
AFLP	Amplified Fragment Length Polymorphism
BAC	Bacterial Artificial Chromosome
CB	Consensus Band
cDNA	Complementary DNA
EST	Expressed Sequence Tag
FISH	Fluorescence *in situ* Hybridization
GSS	Genome Survey Sequence
HICF	High-information Content Fingerprint
Indel	Insertion/Deletion
kb	Thousand Base Pairs
Mb	Million Base Pairs
MYA	Million Years Ago
ORF	Open Reading Frame
PAC	P1-Artificial Derived Chromosome
PCR	Polymerase Chain Reaction
qRT-PCR	Quantitative Reverse Transcriptase PCR
QTL	Quantitative Trait Loci
RFLP	Restriction Fragment Length Polymorphism
RI	Recombinant Inbred

SAS	Simultaneously Amplified Singleton
SSR	Simple Sequence Repeat
STS	Sequence Tagged Site
TE	Transposable Element

CHAPTER I

INTRODUCTION

The grass family Poaceae is a monophyletic family of monocotyledonous flowering plants with ~10,000 species (Kellogg, 2001), which includes the cereal crops, rice, maize, barley, wheat, sorghum, oat, rye, and millet. The cultivated cereal crops provide 60% of the world's food production and a growing share of biofuel. Besides their economic importance, the Poaceae is suitable for genomics study because this family includes many important crops with a vast variation in genome size and diverse native distribution. The genome DNA content of this family exhibits an enormous variability, ranging from ~430 million base pairs (Mb) in diploid rice to ~16,000 Mb in hexaploid wheat (Arumuganathan and Earle, 1991). The independent domestication of rice in both Africa and Asia, sorghum in Africa, maize in America, and wheat in the Near East has also provided an excellent study system in which to explore the genetic complexity of plant adaptation to diverse environments.

This dissertation follows the style and format of Plant Physiology.

ORYZA SATIVA

Rice (*Oryza sativa* [L.], 2n=24) is a C_3 photosynthesis plant, which feeds over half of the global population. It is an ideal model cereal because it has the smallest genome (~340-400 Mb) of the major cereals, is relatively easy to genetically transform, and a vast germplasm collection and dense genetic maps are available.

The importance of rice is reflected in the fact that it was the first crop in which the whole genome for two major subspecies, ssp. *japonica* and ssp. *indica* were sequenced. The sequence of the *japonica* cultivar, Nipponbare, was completed by a multi-national effort, the International Rice Genome Sequencing Project (IRGSP), and represents a map-based clone-by-clone sequencing strategy (Sasaki and Burr, 2000). The draft sequence of the *indica* cultivar was derived by a whole-genome shotgun sequencing approach (Yu et al., 2002; Yu et al., 2005). Although there is a discrepancy in predicted gene numbers between *japonica* and *indica,* the automated annotations indicate a gene complement of ~40,000 to 50,000 genes (Kikuchi et al., 2003; Bennetzen et al., 2004; Yu et al., 2005). In addition to the assembled genomic sequences, public databases currently contain 1,183,556 rice expressed sequence tags (ESTs) (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). A collection of more than 32,000 non-redundant rice full-length complementary DNA (cDNA) sequences is also publicly available (Kikuchi et al., 2003).

All these resources are invaluable not only for rice but also for other cereals. The rice genomic tools can be applied directly to other cereals and the rice genome sequences and annotations can greatly benefit research of other cereals that share an

appreciable degree of synteny with rice. It can also be a great help on uncovering the evolutionary events that shaped the cereal lineage and organizing comparative information for cereals as a single genetic system.

SORGHUM BICOLOR

As the fifth most important cereal crop worldwide (http://faostat.fao.org), sorghum (*Sorghum bicolor* [L.] Moench, 2n=20) not only provides food, feed, and fiber, but also has potential importance in biofuel production (Gnansounou et al., 2005). Due to its excellent tolerance to dry and hot environments, this crop is especially important in arid and semi-arid areas, like Africa. Furthermore, expanding world population coupled with global climatic warming makes sorghum of increasing importance.

As a model of C_4 plants that can increase the efficiency of CO_2 fixation at high temperatures, sorghum is a logical complement to rice and also a suitable model for genomic study. It has a diverse germplasm collection of ~40,000 accessions (Doggett, 1988) and a relatively small diploid genome of ~760-810 Mb (Arumuganathan and Earle, 1991; Price et al., 2005). Sorghum is more closely related to major crops of tropical origin such as maize, sugarcane and pearl millet than rice is, and thus can provide a better roadmap for the genomic study of these crops that have more complex genomes.

Extensive genomic resources have been developed for sorghum (Sorghum Genomics Planning Workshop Participants, 2005). A number of low-density genetic maps have been constructed and cross referenced to other grass species (Whitkus et al., 1992;

Melake Berhan et al., 1993; Chittenden et al., 1994; Pereira et al., 1994; Xu et al., 1994; Dufour et al., 1997; Peng et al., 1999; Bhattramakki et al., 2000; Kong et al., 2000). More recently, Menz et al. (2002) constructed a high-density genetic map in a recombinant inbred (RI) population from a cross of BTx623 and IS3620C. This high-density sorghum map contains 2926 loci, including 2454 amplified fragment length polymorphisms (AFLPs), 203 restriction fragment length polymorphisms (RFLPs) and 136 simple sequence repeats (SSRs), of which 692 of the markers comprise a LOD ≥ 3.0 framework genetic map with an average spacing of approximately 1 cM.

Great efforts have been carried out to construct the sorghum genome physical map. Sorghum was the first plant for which a bacterial artificial chromosome (BAC) library was reported (Woo et al., 1994). The construction of the TAMU-ARS sorghum physical map relied on DNA fingerprinting using polyacrylamide gels to generate contigs of overlapping BAC clones and then AFLP technology was used to locate the contigs on the high-density genetic map as well as to aid in contig merges (Klein et al., 2000). The AFLP analysis was streamlined by incorporating a six-dimensional (6D) BAC pooling strategy that allowed high-throughput identification of BAC clones containing a common DNA marker (Klein et al., 2000). Moreover, since this BAC-based physical map is linked to the genetic map of Menz et al. (2002) it can provide direct access of BAC clones containing genes of agronomic importance that reside within mapped trait loci (Klein et al., 2005).

The architecture of sorghum chromosomes has also been characterized in several studies. A molecular karyotype of the sorghum genome was developed (Kim et al.,

2002) and a uniform sorghum chromosome numbering system was established which has been adopted by the sorghum research community (Kim et al., 2005b). In addition, the molecular cytology of sorghum chromosomes has been analyzed in detail using genetically mapped BAC clones and fluorescence *in situ* hybridization (FISH) technologies (Islam-Faridi et al., 2002; Kim et al., 2005a; Kim et al., 2005c).

Based on the established physical map (Klein et al., 2000), rapid progress has been made towards an integrated genetic, physical and cytogenetic map by using a combination of AFLP, DNA fingerprinting, 6D BAC pooling and FISH technologies. This high resolution integrated genome map will enable efficient map-based isolation of genes, targeted genome sequencing, detailed investigation of genome architecture, comparative analysis with the genomes of other plants and association studies that link DNA marker to important phenotypes.

Approximately 200,000 sorghum ESTs have been collected and revealed ~22,000 unique transcript clusters (Sorghum Genomics Planning Workshop Participants, 2005). Based on these sequences, microarrays and qRT-PCR assays have been carried out to study sorghum gene expression modulated by plant hormones involved in plant protection (Salzman et al., 2005) and osmotic stress (Buchanan et al., 2005). Over 90% of the sorghum genes have been tagged by the collection of ~500,000 methyl filtered sorghum sequences (Bedell et al., 2005). By using a novel direct selection technology, sorghum ESTs can be directly located on the sorghum genome map, allowing up to 75% of the genes encoded in any region of the sorghum genome to be identified (Childs et al., 2001).

Sorghum is known for its recalcitrance to transformation and much effort has been spent to improve this situation. Rapid and reproducible agrobacterium-mediated transformation methods have made it possible for stable sorghum transformation (Zhao et al., 2000; Howe et al., 2006). Improved transformation efficiency, together with progress in transposon tagging (Chopra et al., 1999) and virus-induced gene silencing (Wesley et al., 2001) will dramatically benefit the discovery and analysis of sorghum genes.

Recently, the U.S. Department of Energy Joint Genome Institute Community Sequencing Program announced plans to produce an 8× whole-genome shotgun sequence of the *Sorghum bicolor* genome for public release (http://www.jgi.doe.gov/sequencing/cspseqplans2006.html). The availability of this sorghum genome sequence will benefit not only sorghum, but also other cereals, especially maize and sugarcane that are close relatives to sorghum with more complex genomes.

COMPARATIVE GENOMICS

Comparative genomics is the study of relationships between the genomes of different species or strains to understand the function and evolutionary processes that act on genomes that shaped the modern species. While it is still a young field, comparative genomics is growing rapidly, in part, due to the development of massive databases, the design of efficient query and comparison software tools and ever improving computer technologies.

Although the angiosperm (flowering plants) lineage is believed to be ~200 million years old, the cereals began to diverge from a common ancestor ~55-70 million years ago (MYA) (Doebley et al., 1990). The genome DNA content of the grass family exhibits an enormous variability, ranging from ~430 Mb in diploid rice to ~16,000 Mb in hexaploid wheat (Arumuganathan and Earle, 1991). Despite the vast differences in genome size among the grasses, comparative genetic mapping using common DNA markers has revealed that the relative location of markers and mapped genes show significant conservation (Bennetzen, 2000). The first comparative study was done between sorghum and maize and extensive colinearity was identified (Hulbert et al., 1990). Extensive colinearity was also identified between rice and maize (Ahn and Tanksley, 1993) and rice and the *Triticeae* tribe (Devos and Gale, 1997). The first crop circle detailing the cereals as a single genetic system was published in 1995 for rice, maize, sorghum, sugarcane, foxtail millet, and wheat (Moore et al., 1995), and has since been updated to more precisely delineate syntenic relationships between species of the grass family including rice, oats, maize, sorghum, sugarcane, foxtail millet, wheat, and finger millet (Devos and Gale, 1997; Gale and Devos, 1998).

The traditional way of generating comparative maps is to use common sets of probes with good cross-hybridization ability, primarily cDNAs, across different species that provides limited efficiency and resolution. The recent large-scale EST projects in several grass species have provided a more efficient way to conduct comparative mapping. For example, analysis of a given cereals' ESTs against the rice genome provides the anchor points in rice thereby requiring that the mapping is needed only in

the species from which the ESTs are derived. For genetic regions covered by BAC contigs, a further level of refinement can be obtained by the alignment of low-pass BAC sequence scans with the rice genome sequence (Klein et al., 2003). Efficient query and comparison software tools can also provide another level of accuracy.

Studies examining macrocolinearity (conservation of gene order across regions spanning many megabases) of grass genomes, have suggested that grass genomes can be regarded as single genetic system and information acquired from one grass species can be applied to other grass species (Bennetzen and Freeling, 1993, 1997). This unified genetic system depicted in the crop circle has helped to illustrate the evolutionary relationship between different grass genomes, identify orthologous genes across grass species and facilitate map-based cloning of agronomically important genes, such as the green revolution gene (Peng et al., 1999) and a vernalization gene (Yan et al., 2003).

Over the last several years, comparative genomics in the grasses has progressed from macrocolinearity at the genetic map level toward detailed assessments of microcolinearity at the DNA sequence level. In contrast to map-based efforts, which provide an overview of chromosomal rearrangements that differentiate related species, sequence-based comparisons assess whether gene order, number and orientation have remained conserved within orthologous chromosomal segments. Although grass genomes have large regions of colinear gene order, colinearity has been shown to break down at a finer level that cannot be investigated by standard recombination maps. Numerous exceptions to colinearity including segmental duplication, gene amplification, gene loss, and gene translocation have been detected when orthologous regions of

genomes have been examined at the DNA sequence level (Song et al., 2002). Regions containing fast evolving genes, such as disease resistance genes, are particularly susceptible of breaking down colinearity between related species (Leister et al., 1998). Although a lack of microcolinearity can complicate comparative genetic studies, such sequence rearrangements are a valuable resource in characterizing local genome evolution. Furthermore, the study of microcolinearity is essential to facilitate the map-based cloning of important genes from species with large complex genomes by using the information from the species with the smaller, less complex genome.

Sorghum and maize diverged from rice ~50 MYA and then these two species began to diverge ~16 MYA (Doebley et al., 1990). Compared to rice, the sorghum genome is ~2.2-fold larger in size and has two fewer chromosomes. The increase in DNA content is not uniform across the homeologous sorghum and rice chromosomes, where the pericentromeric heterochromatic regions account for more of the increase in DNA content than the euchromatic portions of the chromosomes (Kim et al., 2005a). The chromosome number difference between sorghum and rice is due to two chromosome fusion events that occurred more than ~20 MYA prior to the divergence of sorghum from maize and *Pennisetum* (Wilson et al., 1999; Kellogg, 2001). Because of these two chromosomal fusion events, sorghum chromosome 1 contains DNA corresponding to rice chromosomes 3 and 10, while sorghum chromosome 2 contains DNA corresponding to rice chromosomes 7 and 9 (Wilson et al., 1999; Kellogg, 2001). All other sorghum and rice chromosomes are homeologous (*Sb*-03:*Os*-01, *Sb*-04:*Os*-02, *Sb*-05:*Os*-11, *Sb*-06:*Os*-04, *Sb*-07:*Os-0*8, *Sb*-08:*Os*-12, *Sb*-09:*Os*-05, *Sb*-10:*Os*-06) (Moore et al., 1995;

Wilson et al., 1999). Moreover, rice and sorghum chromosomes exhibit significant macrocolinearity indicating that sorghum's genome expansion relative to rice is not due to a large-scale genome duplication event (Peng et al., 1999; Wilson et al., 1999; Klein et al., 2003) even though there is strong evidence for segmental and possibly a whole-genome duplication in the common progenitor of sorghum and rice (Paterson et al., 2003; Yu et al., 2005). Furthermore, the sorghum and rice genomes are thought to encode a similar number of genes (http://fungen.botany.uga.edu/Projects/Sorghum/SorghumUnigeneSet.htm; http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml).

SORGHUM CHROMOSOME 3 VS. RICE CHROMOSOME 1

Rice chromosome 1 is the largest (~43 Mb) and first rice chromosome that was sequenced (Sasaki et al., 2002). Based on the latest pseudomolecule release (version 4, http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml) it is estimated that there are ~6100 non-TE (transposable element) gene models located on rice chromosome 1.

Sorghum chromosome 3 is the third largest sorghum chromosome with a size of ~89.9 Mb, of which ~44% (~39.5 Mb) is located in a pericentromeric heterochromatic region. Euchromatic DNA spans ~21.1 Mb of the distal portion of the short arm and ~30.6 Mb of the distal portion of the long arm of sorghum chromosome 3 (Kim et al., 2005a).

Sorghum chromosome 3 is largely colinear with rice chromosome 1 with the exception of one major inversion on the short arm (Ventelon et al., 2001; Klein et al.,

2003). This colinearity will facilitate the approach for constructing a minimal tiling path across the euchromatic arms of sorghum chromosome 3 by allowing for the utilization of the available rice chromosome 1 sequence information.

Several important trait loci, including aluminum tolerance and stay-green, have been mapped to sorghum chromosome 3 (Xu et al., 2000a; Magalhães et al., 2004). Aluminum toxicity is a major limiting factor for plant growth on acidic soils throughout the world. The sorghum aluminum tolerance gene, $Alt_{SB,}$ is the first major aluminum tolerance gene that has been mapped to the genome of a grass species not in the tribe *Triticeae*. It has been mapped near the bottom of sorghum chromosome 3 and is likely to be orthologous to a major aluminum tolerance quantitative trait loci (QTL) on rice chromosome 1 (Magalhães et al., 2004).

Drought is a major constraint on sorghum yield, especially during grain filling. Stay-green is an important component of the post-flowering drought response, which makes sorghum drought resistant during grain filling (Rosenow and Clark, 1981). Under water-limited environments, stay-green genotypes remain green due to a delay in leaf senescence and continue to fill grain. Higher nitrogen content in leaves (Borrell and Hammer, 2000a) and enhanced transpiration efficiency (Borrell et al., 1999) under post-anthesis drought allow these stay-green genotypes to remain photosynthetically active, resulting in more grain yield as compared to senescent genotypes. Also, stay-green genotypes have resistance to charcoal rot (Rosenow, 1983) and lodging (Woodfin et al., 1998), as well as superior ruminant nutritional quality due to higher basal stem sugar content (Duncan et al., 1981) and an increased level of cytokinins (McBee, 1984).

Furthermore, stay-green genotypes can be grown in both irrigated and non-irrigated land because they do not show reduced yield under fully irrigated conditions (Borrell et al., 2000b). Four major stay-green QTL located on three linkage groups were identified using an RI population from BTx642 and Tx7000. Two major QTLs, *Stg1* and *Stg2*, are located on chromosome 3 (Xu et al., 2000b), which explain 20% and 30% of the phenotypic variability, respectively (Sanchez et al., 2002).

Construction of a minimal tiling path across the euchromatic regions of sorghum chromosome 3 utilizing the colinearity between this sorghum chromosome and the nearly complete sequence of rice chromosome 1 will greatly facilitate the molecular cloning of these agronomically and biologically important genes. In addition, examination of microcolinearity of a large orthologous region and analysis of gene expression between sorghum and rice will provide valuable information on cereal genome structure and gene function.

CHAPTER II

CONSTRUCTION OF A MINIMAL TILING PATH ACROSS THE EUCHROMATIC ARMS OF SORGHUM CHROMOSOME 3

INTRODUCTION

Integrated genetic and physical maps are extremely important for map-based gene cloning, comparative genome analysis, and genome sequencing. However, constructing an integrated genetic and physical map can be a difficult task for many plants because of large genome size, repetitive DNA, and polyploidy. Various methods for physical map construction have been developed and one popular method uses restriction enzymes to generate large numbers of fragments from genomic subclones (Brenner and Livak, 1989; Gregory et al., 1997; Marra et al., 1997). Related genomic subclones are identified and assembled into contigs by comparing these fragments that are called DNA fingerprints. However, this method can't link genomic clones directly to genetic maps and the utility is limited due to variation in DNA migration from gel to gel, the presence of repetitive DNAs, unusual distribution of restriction sites, and skewed clone representation. Therefore, a combination of fingerprinting and other methods is required to construct high quality maps of complex genomes. Both PCR- and hybridization-based methods have been used by numerous researchers to combine DNA fingerprinting with genetic

marker content mapping for the construction of integrated genome maps (Marra et al., 1997; Cao et al., 1999; Vollrath and Jaramillo-Babb, 1999; Zhu et al., 1999).

Klein et al. (2000) published a reliable, low cost and highly efficient method for the construction of an integrated sorghum genetic and physical map. This high-throughput method relied on a combination of DNA fingerprinting, BAC DNA pooling, and AFLP-based contig assembly and mapping. AFLPs have been used effectively as a high-throughput genetic marker system (Alonso-Blanco et al., 1998; Qi et al., 1998; Boivin et al., 1999; Vuylsteke et al., 1999; Young et al., 1999). In the method of Klein et al. (2000), AFLP technology was combined with a six-dimensional BAC DNA pooling strategy. In this pooling strategy, BAC clones representing 4-5× genome equivalents were pooled on six coordinate axes to create 184 pools each containing either 796 or 1024 BAC clones. The DNA from each pool was extracted and used in PCR-based assays to provide for the efficient identification of BAC clones harboring AFLP markers as well as other PCR-based marker types (i.e., SSR, STS, and Indel).

In 2003, a high quality framework genetic and physical map of sorghum chromosome 3 was constructed and aligned to rice chromosome 1 following sequence scanning of subclones from 118 BAC clones linked to 109 genetic loci along the length of the chromosome (Klein et al., 2003). In addition, a newly devised method of four-color high-information content fingerprint (HICF) analysis (Ding et al., 2001; Luo et al., 2003) in conjunction with EST-STS marker content mapping based on alignment to the rice genome sequence was evaluated for the ability to close gaps in the framework physical map thereby creating large contiguous contigs (Klein et al., 2003).

In the modified HICF method, BAC DNA is simultaneously digested with four different 6-base (*Eco*RI, *Bam*HI, *Xho*I, and *Xba*I) and one 4-base (*Hae*III) recognition restriction enzymes. Each of the four 6-base recognition enzymes produces a different 1-base 5' overhang that is subsequently filled in using one of the four fluorescent dideoxy terminators in the SNaPshot™ labeling kit. Each product is therefore characterized by both the restriction enzyme producing the fragment and the size of the fragment. The HICF method produces four times more information than standard single restriction enzyme fingerprinting, and has been shown to significantly increase the accuracy and efficiency of detecting BAC clones with shared fragments (Ding et al., 2001; Luo et al., 2003; Nelson et al., 2005). However, there are some inherent limitations of the HICF method. HICF fragments need to be quite small, currently 500bp or less, and double peaks can't be scored. In addition, HICF projects tend to contain many Q, or questionable clones. These arise because of errors in the band files due to both spurious bands and missing bands. A high number of Q clones can lead to false contig merges. This problem can be resolved by running a program called DQer in the FPC program which will reorder HICF bands at a stricter cutoff (Nelson et al., 2005). Even with these limitations, the HICF approach is becoming increasingly popular requiring little manual intervention, while providing excellent precision in sizing contigs and greater sensitivity in forming contigs (Nelson et al., 2005).

Sorghum chromosome 3 is largely colinear with rice chromosome 1 with the exception of one major inversion on the short arm (Ventelon et al., 2001; Klein et al., 2003). This colinearity can facilitate the approach for constructing a genetic and

physical map of sorghum chromosome 3 by allowing for the utilization of the available rice chromosome 1 sequence information. An EST-STS PCR-based approach was developed that utilized the aligned rice genome sequence to aid in gap filling in the sorghum physical map (Klein et al., 2003). This method relies on identifying low- or single-copy sorghum ESTs that are homeologous to genes from the rice chromosome 1 pseudomolecule that reside in gaps between two sorghum contigs and using this information to design EST-STS primers for subsequent screening of the BAC DNA pools to identify those BAC clones which contain that particular EST. The BAC clones identified using this method are then fingerprinted by HICF, along with the previously mapped BAC clones flanking the gap in the physical map, to determine their order, degree of overlap, and potential for filling gaps in the sorghum physical map. Using this strategy, an ~1.6 Mb contig across a region of the long arm of sorghum chromosome 3 was rapidly and efficiently constructed demonstrating the utility of this method for constructing a minimal tiling path across the euchromatic region of any chromosome of interest (Klein et al., 2003). This strategy is similar to the approach utilized for the construction of a sequence-ready physical map of the mouse genome (Kim et al., 2001). In the present study, a minimum tiling path across the euchromatic arms of sorghum chromosome 3 was constructed using this strategy since it is efficient, cost-effective, requires minimal technical skill, and can easily be implemented at high throughput.

Constructing a minimal tiling path across the euchromatic arms of sorghum chromosome 3 will yield additional information about the relationships between rice and sorghum genomic segments, provide a well-characterized clone set for sequencing the

euchromatic portion of the sorghum genome, provide complete contigs across *Stg1*, *Stg2* and Alt_{SB} for use in fine-mapping and eventual map-base cloning, and ultimately enhance our understanding of cereal genome structure and evolution.

MATERIALS AND METHODS

EST-STS PCR-based approach to screen BAC pools

The minimum tiling path of rice chromosome 1 was used as the roadmap to construct the minimal tilling path of the euchromatic arms of sorghum chromosome 3. The rice chromosome tiling path was obtained from the Rice Genome Research Program (http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/stattable.pl?chr=1&lab=RGP). BLASTN analysis was used to identify sorghum ESTs homeologous to genes encoded by the individual rice PAC (P1-artificial derived chromosome) clones. Sorghum ESTs that appeared to be single- or low-copy from this analysis were further screened by BLASTN to the GenBank non-redundant database to ensure that each sorghum EST exhibited strong sequence similarity only with the appropriate BAC clones on rice chromosome 1. Sorghum ESTs exhibiting strong sequence similarity to the rice chromosome 1 minimum tiling path with an expect value greater than e^{-20} were selected for primer design. EST-STS primers were designed for every 50-150 kb in which a gap resided within the current physical map of sorghum chromosome 3. Primers to sorghum ESTs were designed using the PrimerQuest™ software program and were obtained from Integrated DNA Technologies (Coralville, IA, USA). PCR-based screening of BTx623 and

IS3620C BAC DNA pools was performed in 10 µl reactions containing 1× Perkin-Elmer buffer II (Applied Biosystems, Foster City, CA, USA), 2.5 mm $MgCl_2$, 200 µm dNTPs, 0.4 U of AmpliTaq polymerase, 20 pmol of each primer, and 5 ng of pooled BAC DNA or genomic DNA (BTx623 or IS3620C). Amplification conditions were as previously described (Klein et al., 1998). Amplification products were electrophoresed in 2% agarose gels and the products visualized following staining with SYBR Gold (Molecular Probes Inc., Eugene, OR, USA). BAC DNA pools containing signals for a given sorghum EST were recorded, and the individual BAC clones containing the EST were identified using a Unix script written in the Perl programming language (Klein et al., 2000; Klein et al., 2003). After the first round of contig construction, gaps in the maps that persisted were targeted during a second round of EST-STS screening as described above.

AFLP screening of BAC pools

AFLP template, *Eco*RI/*Mse*I and *Pst*I/*Mse*I, was prepared from the IS3620C BAC DNA pools as described previously (Klein et al., 2000; Menz et al., 2002). AFLP pre-amplification (0/+1 primers) and selective amplification (+3/+3 primers) reactions of BAC pool template were performed as described (Klein et al., 2000; Menz et al., 2002), using the 192 AFLP primer combinations previously used for the development of the high-density AFLP-based genetic map of sorghum (Menz et al., 2002). Amplification products were analyzed as described using a LI-COR dual-dye DNA sequencing system (LI-COR, Inc, Lincoln, NE, USA) (Klein et al., 2000; Menz et al., 2002). Individual

BAC clones harboring AFLP genetic markers were identified from the BAC DNA pools using the Perl script mentioned above (Klein et al., 2000; Menz et al., 2002). Screening the IS3620C BAC pools with 32 of the 192 +3/+3 primer combinations as well as specific targeted screening of AFLP markers previously mapped to sorghum chromosome 3 was carried out in the present research.

HICF fingerprinting of identified BAC clones

A newly devised method of four-color HICF analysis was used on those BAC clones identified as positive from the EST-STS and AFLP screening described above for contig construction (Luo et al., 2003). BAC DNA was isolated in deep-well plate format as previously described, except for the addition of RNaseA (100 µg/ml) to the initial re-suspension buffer (Klein et al., 1998). Following purification, BAC DNA pellets were re-suspended in 45 µl sterile water and quantified by fluorimetry, and the DNA concentration was adjusted to 50-75 ng/ml with sterile water. BAC DNA was restricted by mixing 42 µl of purified DNA with 9 µl of enzyme cocktail consisting of 5 µl of 10× NEBuffer 2 (New England Biolabs, Beverly, MA, USA), 0.5 µl of BSA (100 mg/ml), 1 µl of RNaseA (0.5 µg/ml), 1 µl of 1% β-mercaptoethanol, 5 U of *Bam*HI, 5 U of *Eco*RI, 5 U of *Xba*I, 5 U of *Xho*I, and 5 U of *Hae*III. The reactions were incubated for 3 h at 37°C. Following restriction enzyme digestion, 10 µl of labeling cocktail containing 2 µl of 10× NEBuffer 2, 2.5 µl of Tris (pH 9.0), 1 µl SNaPshot™ Multiplex Ready Reaction Mix (Applied Biosystems), and 4.5 µl of sterile water was added to each sample, and the reactions were incubated at 65°C for 1 h. Reactions were precipitated by the addition of

5.0 µl of 2.5 M sodium acetate and 100 µl of chilled 95% ethanol (− 80°C, 15 min). Labeled fragments were collected by centrifugation at 1920 g for 30 min, washed once with 70% ethanol, and air-dried. Pellets were re-suspended in 9.75 µl of Hi-Di formamide containing 0.25 µl of GeneScan™-500 Liz size standard (Applied Biosystems). Labeled fragments were separated on an ABI3700 sequencer using the SNP2_POP5 module with modifications. Electrophoresis was performed at 6000 V at a run temperature of 50°C for 90 min. Collected data was analyzed with GeneScan™ version 3.7 Fragment Analysis Software (Applied Biosystems). The minimum peak detection threshold was individually set for each lane and for each fluorescent dye. Data was manually edited using Genotyper™ version 3.7 Fragment Analysis Software (Applied Biosystems), and a table consisting of a list of fragment sizes and dye colors for each BAC clone was exported in a tabular format. Conversion of the Genotyper™ output table into a band file for input into FPC V8.2 was accomplished using a script written in the Perl programming language. Automated contig assembly was performed using FPC V8.2 based on BAC overlap relationship (Soderlund et al., 1997) followed by manual inspection to merge contigs or break falsely joined contigs based on all other available information including BAC sequence scan data and genetic and EST-STS marker data.

RESULTS AND DISCUSSION

EST-STS PCR-based approach to screen BAC pools

The previous alignment of sorghum chromosome 3 with rice chromosome 1 allowed for the prediction of sorghum gene locations based on the colinearity with most rice genes (Klein et al., 2003). The identification of sorghum genes that are potentially located within gaps on the sorghum physical map provides the necessary information to identify those BAC clones that encode these genes. Positive clones that are identified can then be fingerprinted and used for physical map construction.

In total 612 single- or low-copy sorghum ESTs with homologs on rice chromosome 1 PAC clones were selected for EST-STS screening of BAC pools based on their potential to fill gaps in the sorghum physical map (Table I). Four hundred and sixty-six EST-STS primer sets produced strong PCR signals in the BAC pools that could be easily analyzed, of which 447 primer sets identified BTx623-derived BAC clones and 452 identified IS3620C-derived BAC clones, while the other primers either amplified a band/bands from every DNA pool or failed to amplify a product even from sorghum genomic DNA.

Figure 1 shows a representative agarose gel image following screening of the BTx623 BAC pools with primers designed from EST CN141545 (ESTs are denoted by their GenBank accession number). One BTx623 clone, sbb6537, was resolved as positive in this analysis.

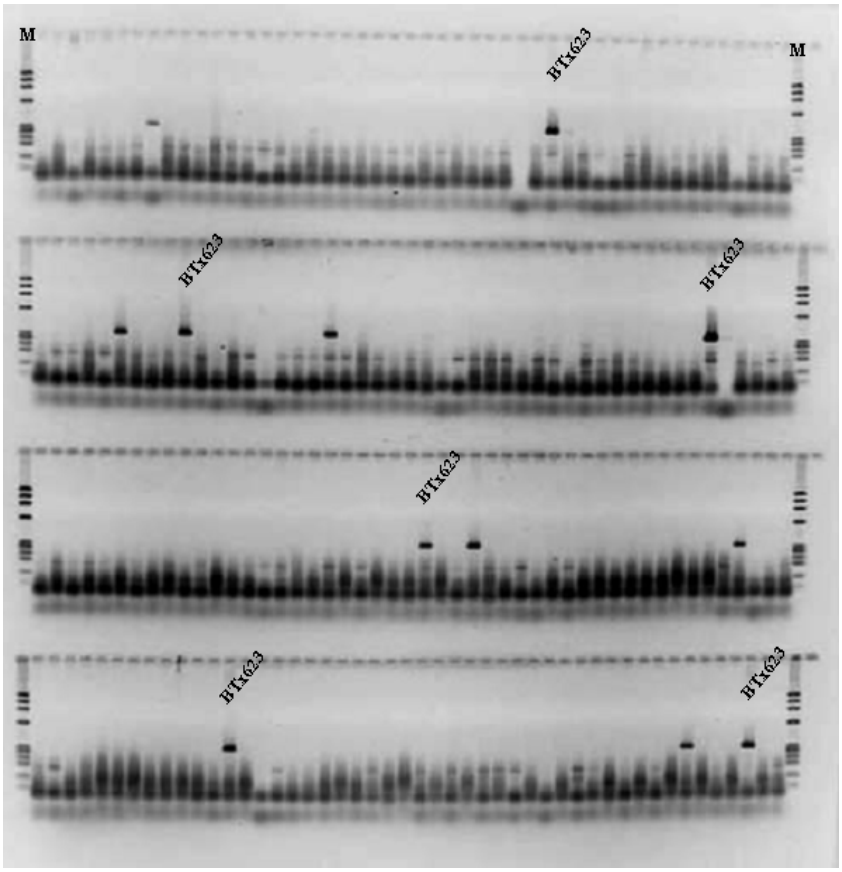

Figure 1. Gel image of EST-STS PCR-based BAC pool screening. BTx623 BAC DNA pools were screened with EST primers designed from EST CN141545. One band is detected in each pool and finally resolved as BTx623 clone sbb6537 by the Unix Perl script. BTx623 genomic DNA was used to separate each pool type and prove the validity of the reaction as shown in the image. PhiX 174 DNA *Hae*III digested molecular weight markers were loaded on the side of the gel and marked as M.

Table I. Summary of EST-STS primers designed for sorghum BAC pool screening

	Short Arm	Heterochromatin	Long Arm	Total
Number of sorghum ESTs selected for EST-STS primer design	191	45	376	612
Number of primers producing a strong and analyzable band in BTx623 pools	136	35	276	447
Number of primers producing a strong and analyzable band in IS3620C pools	141	34	277	452
Total number of primers producing a strong and analyzable band	145	36	285	466

In total, 2,080 sorghum BAC clones were identified by the EST-STS primer screening from both BTx623 and IS3620C BAC pools. One thousand and ninety-three clones were identified from the BTx623-derived BAC pools and 987 clones were identified from the IS3620C-derived BAC pools (Table II). On average, each primer set identified ~2.45 BTx623 BAC clones and ~2.18 IS3620C BAC clones following screening of their respective pools. However, 34 EST-STS primer sets did not identify any BTx623 BAC clones and 13 primer sets did not identify any IS3620C BAC clones. In addition, one primer set did not identify any clones in either the BTx623 and IS3620 BAC pools, while this primer set did generate a strong signal in genomic DNA (data not shown). These results indicate ~92% coverage of the sorghum genome by the BTx623 BAC pools and ~97% coverage of the sorghum genome by the IS3620C BAC pools. The sorghum BAC pools used in this research each provide ~4× coverage of the sorghum genome. These results are in good agreement with previous results in which the BTx623 BAC DNA pools were screened for over 300 SSR, STS and AFLP markers and nearly 98% coverage was observed (Klein et al., 2000). When the results from the BTx623 and IS3620C BAC pools are combined together, the genome coverage increases to ~99.8%.

Table II. Number of BAC clones identified by EST-STS screening of BTx623 and IS3620C pools

	Short Arm	Heterochromatin	Long Arm	Total
BAC clones identified from BTx623 pools	345	69	670	1093
Number of ESTs not present in BTx623 pools	8	0	26	34
Average number of BTx623 BAC clones identified per EST-STS primer	2.6	2.76	2.34	2.45
BAC clones identified from IS3620C pools	296	51	640	987
Number of ESTs not present in IS3620C pools	16	5	34	55
Average number of IS3620C BAC clones identified per EST-STS primer	2.1	2.13	2.23	2.18

Sorghum ESTs were used to design primers for pool screening. Currently, more than 200,000 sorghum ESTs have been sequenced and assembled into ~22,000 unique transcripts, representing ~40-50% of the total gene space (Sorghum Genomics Planning Workshop Participants, 2005). In addition, ~500,000 methyl-filtered sequences from sorghum have been collected and assembled into contigs providing an estimated 1× coverage of the methyl-filtered gene space (Bedell et al., 2005). These resources provide for significant coverage of the gene space across the sorghum genome for use in physical map construction. Because only low- or single-copy EST sequences are selected for primer design, this method reduces the chances of screening the BAC pools with primers designed against repetitive genes or large gene family members. Screening the pools with primers that amplify large gene family members or repetitive sequences results in the identification of numerous BAC clones, many of which will map to other locations within the sorghum genome. This can be a significant problem during physical map construction resulting in incorrect merging of contigs during manual inspection. On average, each primer pair identified ~2.3 BAC clones. By employing the reciprocal BLAST analysis step prior to primer design, there were very few chances of identifying more than 8 BAC clones with a given primer set which has been shown to result in an increase in the number of false-positive clones leading to problems in accurate contig construction (Klein et al., 2000). Therefore, the accuracy and order of the map can be greatly enhanced by the current method in contrast to other methods that rely on a hybridization-based approach using small overgo probes (Draye et al., 2001; Bowers et al., 2005).

The sorghum ESTs were not evenly distributed across the chromosome, most of them were located in the euchromatic regions and very few were found in the heterochromatin. Only 45 low- or single-copy ESTs were identified for primer design in the heterochromatin although this block of DNA represents nearly 44% of sorghum chromosome 3 (~39 Mb), while 567 ESTs were easily identified for primer design within the ~51.7 Mb euchromatic region (Kim et al., 2005a). Even in the euchromatic regions, the ESTs were not evenly distributed (Data not shown). It is believed that most of the existing gaps in the minimal tiling path could be closed if more ESTs existed in the gap regions.

Rice is a model cereal and its available sequence information can benefit other cereals, like sorghum and maize, which share an appreciable degree of synteny. Sorghum chromosome 3 is largely colinear with rice chromosome 1 with the exception of one major inversion on the short arm (Ventelon et al., 2001; Klein et al., 2003). Based on the available rice chromosome 1 sequence information, BLASTN analysis can easily identify sorghum ESTs homeologous to genes encoded by individual rice PAC clones. The strategy utilized here to construct an extended BAC contig using the rice genomic sequence is similar to the approach utilized for the construction of a sequence-ready physical map of the mouse genome (Thomas et al., 2000; Kim et al., 2001).

A six-dimensional BAC DNA pooling strategy was used for the PCR-based approach to identify colinear sorghum BAC clones. This pooling strategy was designed to allow 4-5× genome equivalents of DNA to be screened for the presence of BAC clones containing the same marker, such as STS, SSR, or AFLP markers, in a single

step. In the current work, the 6D BAC pools were combined with EST-STS PCR and AFLP technology to rapidly and efficiently identify the markers of interest. The average number of positive BAC clones identified per marker was less than what was expected for a library containing 4-5× genome equivalents. This may be caused when a clone is occluded within the stack of BAC clones or when a signal in one or more pool types is missing since our pooling strategy requires that a PCR signal be detected in all six unique pool types to be considered a true positive. However, these potentially false-negative clones can be marked and individually confirmed for the presence of the marker. The present results indicate that 6D BAC pooling combined with PCR-based screening is efficient, cost-effective, requires minimal technical skill, and can easily be implemented at high throughput. Other strategies utilizing superpools and subpools in PCR-based screening approaches have also been used successfully to identify individual positive clones, but these methods were not compatible with our desire to screen a redundant library representing a large genome simultaneously in a single step (Green and Olson, 1990; Asakawa et al., 1997).

AFLP screening of BAC DNA pools

AFLP screening of BTx623 BAC pools had been previously completed utilizing 192 +3/+3 primer combinations (Klein et al., 2000). AFLP screening of the IS3620C BAC library pools was carried out to link the IS3620C BAC clones to the high-density genetic map in a similar fashion. Analysis of 32 of the 192 +3/+3 primer combinations was carried out as a part of this research and the remaining 160 primer combinations were

done by other members of the lab. Data was analyzed from all 32 +3/+3 primer combinations using Bionumerics software and the clones positive for a given AFLP identified and marked accordingly in the physical map. In addition, targeted screening of AFLP markers previously located on the sorghum chromosome 3 high-density genetic map (Menz et al., 2002) was carried out as part of this work to link as many IS3620C BAC clones from the physical map to the chromosome 3 genetic map as possible. A representative example of an AFLP gel image is displayed in Figure 2. After the analyses of the amplification products, individual BAC clones harboring AFLP genetic markers were identified from the BAC DNA pools using a perl script and the information added to the FPC database. In the present study, 94 AFLP markers (59 located within the euchromatic arms and 35 located within the heterochromatin) were scored for their presence in the IS3620C BAC DNA pools providing additional linkage between the genetic and physical maps. In total, the minimal tiling path of sorghum chromosome 3 contains 290 genetic markers, which includes the AFLP markers added by this and previous research (supplemental Table X).

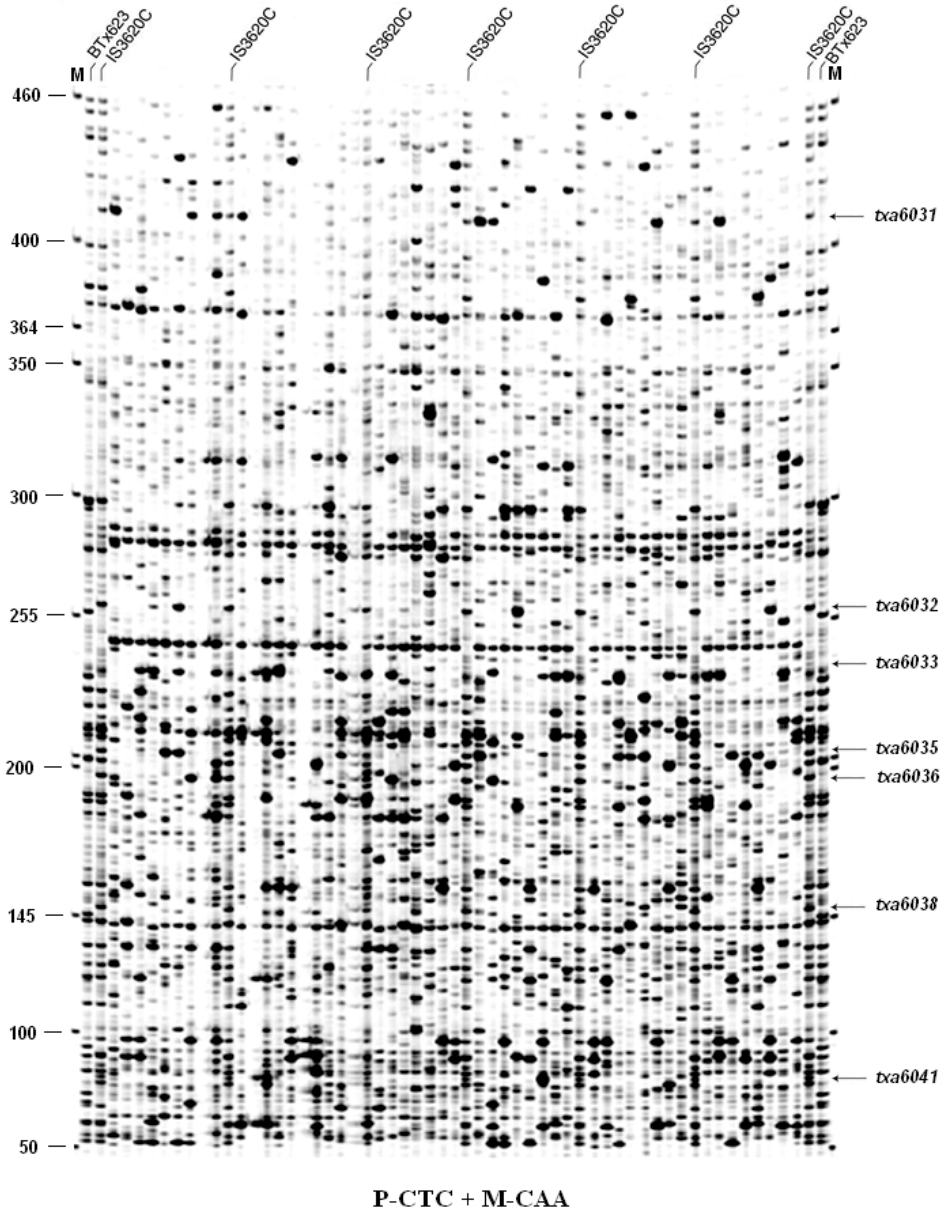

Figure 2. PCR-based screening of BAC DNA pools using AFLP technology. AFLP templates were prepared from IS3620C BAC pools and selectively amplified with fluorescent-labeled *Pst*I + CTC and *Mse*I + CAA primers. Labeled products were analyzed on a LI-COR DNA sequencer. AFLP templates from genomic DNAs, BTx623 and IS3620C, were run as controls and are indicated above the respective lanes. Arrows to the right of the gel show the polymorphic AFLP markers. Fluorescent-labeled molecular weight markers (LI-COR) were run in lanes marked M and their sizes, in bp, are shown to the left of the gel.

The AFLP technique has been used effectively as a high-throughput genetic marker system and has the advantages of efficiency and low cost over other traditional marker systems (Alonso-Blanco et al., 1998; Qi et al., 1998; Boivin et al., 1999; Vuylsteke et al., 1999; Young et al., 1999). In this research, AFLP technology was combined with the BAC DNA pooling strategy to allow overlapping BAC clones to be identified while simultaneously linking the genetic and physical maps. In addition to identifying and scoring for AFLP bands polymorphic in the mapping parents used to construct the BAC libraries, numerous monomorphic bands (i.e., non-polymorphic between the mapping parents) were also identified and scored through the BAC DNA pools. These unique amplified fragments, termed SAS-DNAs (simultaneously amplified singleton DNAs) (Klein et al., 2000) are ideal for identifying BAC clones that contain the same unique PCR-amplifiable DNA sequence although they do not provide links to a genetic map. Overall, the combined approach of AFLP technology and BAC DNA pooling provides a low cost, efficient way for building high quality integrated genetic and physical genome maps.

Minimal tiling path across the euchromatic arms of sorghum chromosome 3

BTX623 and IS3620C BAC clones identified from the EST-STS and AFLP screening were subjected to HICF analysis (Ding et al., 2001; Luo et al., 2003). The HICF method was used to fingerprint all clones positive for the markers analyzed as well as all other BAC clones located within the same contig from a previous fingerprinting project using a polyacrylamide-based method (Klein et al., 2000). Automated contig

assembly was performed initially at a tolerance of 4 and a cutoff of e^{-45} and then the cutoff was successively lowered down to e^{-25}. Following automated contig assembly manual addition of singleton clones and merging of contigs was performed using both clone overlap information as well as marker information to produce 23 contigs that represented the physical map of the euchromatic arms of sorghum chromosome 3 (Table III). There were instances were contigs likely should merge but the overlap of shared bands was not sufficient to reliably perform the merge. For example, contig4076 and contig4611 share two common markers at their ends and contig4611 and contig94 share a common marker at their ends providing strong support that these three contigs should in fact be merged into a single contig. The ends of these contigs contain BAC clones from the two different parental genotypes and thus generated different fingerprinting patterns that can't be joined together by the FPC program. For the current analysis manual merges between contigs was not performed in FPC if the end clones did not show any significant shared fragments between them even if they shared common markers. However, these three contigs are treated as a single contig for size estimation to show the fact that this region is most likely fully covered by BAC clones (Table III).

In this newly devised HICF method (Ding et al., 2001; Luo et al., 2003), BAC DNA is simultaneously digested with four different 6-base and one 4-base recognition restriction enzymes, in which each of the four 6-base recognition enzymes produces a different 1-base 5' overhang that is subsequently labeled by one of the four fluorescent dideoxy terminators. Since each product is therefore characterized by both the restriction enzyme that produced the fragment and the size of the fragment, the HICF

method produces four times more information than standard single restriction enzyme fingerprinting, resulting in a significant increase in the accuracy and efficiency of detecting BAC clones with shared fragments. It has been estimated that HICF requires less than 20% overlap to construct contigs and requires 10-fold fewer clones to be fingerprinted than a system requiring 80% overlap to achieve the same level of closure (Ding et al., 2001). The HICF approach is becoming increasingly popular requiring little manual intervention, while providing excellent precision in sizing contigs and greater sensitivity in forming contigs (Nelson et al., 2005).

A screen shot of the FPC view of contig498 is shown in Figure 3. This contig is one of the smallest constructed containing 35 BAC clones and 19 markers. The contig is 473 CB units, which equates of ~625 kb. Markers shown above the contig include genetic markers (denoted with the chromosome number following the name), EST-STS markers and SAS markers. BAC remarks are shown below the contig map.

Table III. Sorghum BAC contigs representing the sorghum chromosome 3 minimal tiling path. Contigs are listed in the order in which they align to the genetic map

Contig Name	Number of Clones (Buried)	Number of Markers	Length in CB[a]	Length in kb[b]
457	113 (22)	53	1717	2270
8642	196 (50)	90	2923	3864
498	35 (9)	19	473	625
30	4 (0)	4	159	210
8735	53 (19)	2	874	1155
23	165 (0)	49	2241	2962
516	117 (23)	49	2059	2722
301	12 (4)	2	392	518
9539	6 (1)	12	254	336
8664	26 (2)	10	438	579
1232	69 (13)	25	1127	1490
283	25 (3)	1	738	976
9058	8 (0)	1	223	295
1739	190 (43)	57	2878	3805
608	37 (8)	11	927	1225

Table III. Continued

Contig Name	Number of Clones (Buried)	Number of Markers	Length in CB[a]	Length in kb[b]
508	122 (0)	31	2086	2758
77	448 (102)	96	6014	7951
10664	2 (0)	1	181	239
385	220 (38)	80	3901	5157
4076	80 (14)	26		
4611	97 (25)	48	3513	4644
94	134 (27)	86		
26	935 (0)	357	10422	13778
Total	3094(403)	1110	43540	57560

[a]Consensus Band, The ordering of clones and their fragments is called a Consensus Bands map. The coordinate system used in the contig display is in CB units: each distinct band is one unit of measurement (1 CB).

[b]HICF analysis of 15 fully sequenced sorghum BAC clones was performed and the average number of ~75- to 500-bp DNA bands detected by HICF per megabase pair of DNA was calculated. This analysis indicated that 1 CB is equivalent to ~1.322 kb (Kim et al., 2005a).

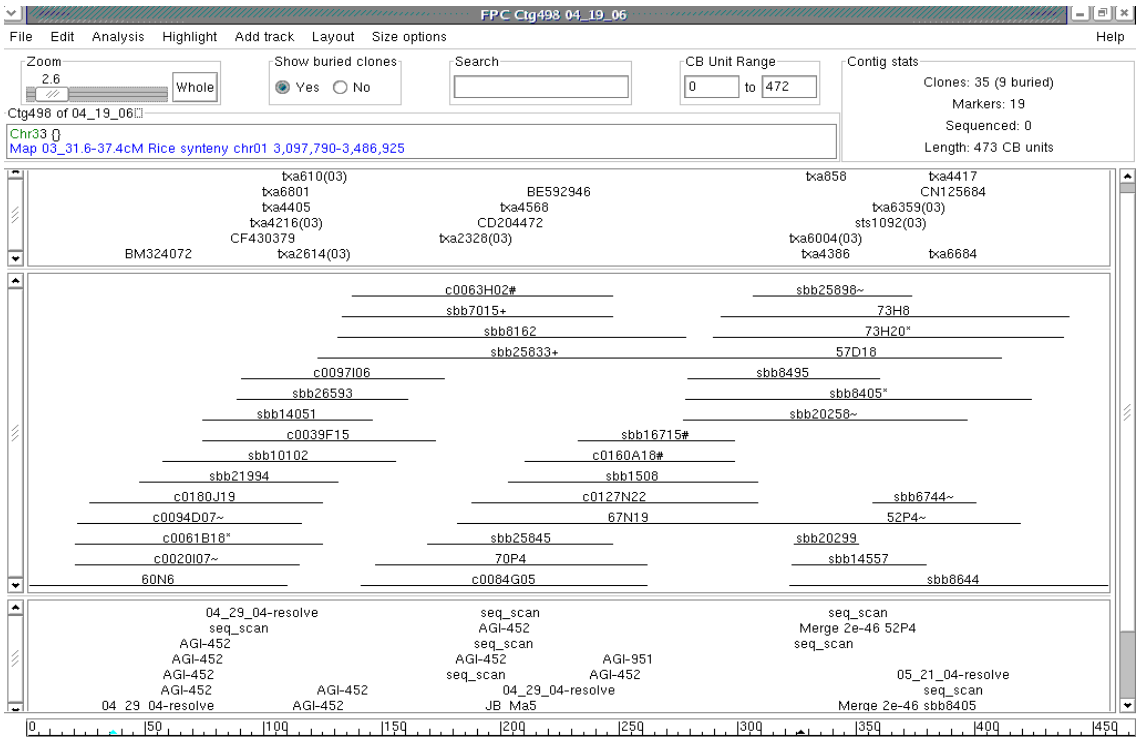

Figure 3. FPC view of sorghum contig498. A screen-shot of an opened FPC window is shown. Thirty-five sorghum BAC clones are shown in this contig with a length of 473 CB units or ~625 kb. Six genetic markers, 5 EST-STS markers and 7 SAS markers are shown above the contig.

In FPC, the length of a contig is measured by the number of consensus bands (CB) that are used to calculate the similarity and relatedness of overlapping clones. It has been determined that 1 CB is equivalent to ~1.322 kb based on HICF analysis of 15 fully sequenced sorghum BAC clones where the average number of ~75- to 500-bp DNA bands detected by HICF per megabase pair of DNA was calculated (Kim et al., 2005a). Using this information, we can transform the contig length from CB to physical length in kb (Kim et al., 2005a). Based on this transformation, the size of the largest contig is 13,778 kb, contains 935 BAC clones, and covers half of the size of the euchromatic region of the long arm. The size of the smallest contig is 210 kb, which contains only 4 BAC clones. The average size of the contigs spanning the euchromatic arms of sorghum chromosome 3 is calculated to be 2.50 Mb. These 23 contigs represent a total of 57.56 Mb of the euchromatic regions of sorghum chromosome 3, of which 25.79 Mb represents the short arm and 31.77 Mb represents the long arm.

Kim et al. (2005a) used FISH analysis to estimate that sorghum chromosome 3 contains a total of 89.9 Mb, of which 51.7 Mb is from the euchromatic regions and the remaining 38.2 Mb resides in the heterochromatin. The present estimates of the DNA represented in the sorghum chromosome 3 minimal tiling path (both euchromatin and heterochromatin) are likely an overestimate of the actual DNA content represented in the minimal tiling path due to the fact that the current HICF map contains BAC clones from two highly divergent sorghum genotypes (BTx623 and IS3620C), and to the inherent limitations associated with BAC DNA fingerprinting (Nelson et al., 2005). In any case, even if the present results are overestimating the DNA content by 10-15% (57.56 Mb in

euchromatin in present study vs. 51.7 Mb based on FISH analysis), the present results suggest that the minimal tiling path represents nearly 100% of the DNA from the euchromatic arms.

Twenty-two gaps existed between the contigs constructed across the euchromatic arms of sorghum chromosome 3. Gap size was estimated based on the size of the colinear rice region except the gap between the two contigs of the short arm that are located at the boundaries of the major inversion. The size of each sorghum gap can be calculated based on the ratio of DNA in the colinear regions of rice and sorghum (Kim et al., 2005a). For this analysis, the EST-STS markers closest to the ends of each contig at a gap were determined, and then the homeologous rice region between these 2 EST-STS markers was determined for size calculation. Where the sorghum contig extended out beyond the EST-STS marker, the size of the extension was determined and then deducted from the final gap size. The largest gap size is 380 kb and some gaps were actually estimated to be 0 in cases where a sorghum BAC within the contig extended beyond the EST-STS marker and its consensus band length was longer than the estimated gap size (data not shown). This may suggest that some of the gaps do not exist but the overlap between the end BAC clones is not significant enough for them to be joined by the FPC program. Even if the gaps do exist, it is likely that these gaps can be easily filled using other methods such as BAC-end sequencing, since only 1-3 BAC clones will be needed to cover the estimated ~100-400 kb distance that remains. Alternatively, once the 8× whole genome shotgun sequence assemblies are released, these gaps should readily be closed. It is possible, however, that some of the gaps may

be larger than estimated. The gap size was estimated based on the ratio of DNA in the colinear regions of rice and sorghum. These colinear regions normally cover a long distance and the estimated ratio may not necessarily represent the specific small region where the gaps exist. The fact that these gaps weren't closed after 2 rounds of EST-STS primer screening may indicate local rearrangements which could have expanded the size of the sorghum genome relative to rice and cause a loss of local colinearity.

Constructing a physical map across the heterochromatic region will be much more difficult due to its high content of repetitive DNA and lack of gene density. Efforts were also undertaken to construct contigs within the heterochromatic region of sorghum chromosome 3 (data not shown). Although gene density within the heterochromatin is much lower than that found in the euchromatic arms (1 gene model/34.5 kb vs. 1 gene model/12.3 kb), estimates indicate that as much as 30% of the gene space in sorghum resides in the heterochromatin (Kim et al., 2005a). A limited number of EST-STS primers could be designed for the heterochromatic region as many rice PAC clones aligned to this region did not show significant homology to any sorghum EST-STS in the database and others were highly repetitive and thus eliminated from further analysis. In total, 36 EST-STS primers designed for the heterochromatic region produced strong and analyzable PCR signals in the BAC pools. Four contigs contained both EST-STS and AFLP markers confirming their location in the heterochromatic region and alignment to the rice chromosome 1 pseudomolecule. Additionally, two other contigs contained blocks of colinear EST-STS markers (3 EST-STS markers in one contig and 5 EST-STS markers in the second contig), however, these two contigs did not contain any

genetic markers and therefore could not be genetically linked to the region. Most of the contigs within the heterochromatic region contain only AFLP markers linking them to the genetic map and these contigs are not aligned to the rice genome sequence. A total of 21 contigs are associated with the heterochromatic region based on analysis of 35 AFLP markers mapped to this region (supplemental Table X). Most of these contigs are quite small with an average size of 789 kb, however, the total distance covered within the heterochromatin was 16.97 Mb, or nearly 45%, of the estimated 38.2 Mb total size (Kim et al., 2005a). All these BAC clones linked to the heterochromatin could serve as seed BACs to construct a minimal tiling path of the pericentromeric heterochromatic block and gaps could possibly be closed by the sequence-tagged connector or chromosomal walking methods that were used successfully in rice (Wu et al., 2004).

CHAPTER III

MACROCOLINEARITY BETWEEN SORGHUM CHROMOSOME 3 AND RICE CHROMOSOME 1

INTRODUCTION

Despite the vast differences in genome size and millions of years evolutionary divergence among the grasses, comparative genetic mapping using common DNA markers has revealed that the relative locations of markers and mapped genes show significant conservation among the cereals (Bennetzen, 2000). This phenomenon was first summarized as the 'crop circle diagram' (Moore et al., 1995) and has since been updated to more precisely delineate syntenic relationships between species of the grass family including rice, oats, maize, sorghum, sugarcane, foxtail millet, wheat, and finger millet (Devos and Gale, 1997; Gale and Devos, 1998).

The traditional way of generating comparative maps is to use common sets of probes with good cross-hybridization ability, mostly cDNAs, across different species. The order of these cDNA markers across species is normally conserved across regions of the genome spanning many megabases. Although this type of comparative analysis of the grass genomes provides only limited efficiency and resolution, it has been instrumental in illustrating the evolutionary relationships between different grass genomes, identifying orthologous genes across grass species, and facilitating map-based cloning of

agronomically important genes, such as the green revolution gene (Peng et al., 1999) and a vernalization gene (Yan et al., 2003). The recent large-scale EST projects in several grass species have provided a more efficient way to conduct comparative mapping across species. When combined with *in silico* analysis of a reference genome sequence (rice genome for example), ESTs need only be mapped in the species in which they were derived.

Sorghum diverged from rice ~50 MYA (Doebley et al., 1990) and their chromosomes exhibit significant macrocolinearity (Moore et al., 1995; Wilson et al., 1999). Sorghum chromosome 3 is largely colinear with rice chromosome 1 with the exception of one major inversion on the short arm (Ventelon et al., 2001; Klein et al., 2003). In this chapter, the macrocolinearity between sorghum chromosome 3 and rice chromosome 1 will be examined using the information derived from the EST-STS mapping work outlined in Chapter II.

RESULTS AND DISCUSSION

Analysis of the macrocolinearity between sorghum chromosome 3 and rice chromosome 1 based on mapped ESTs

From all of the EST-STS markers used to construct the minimal tiling path detailed in Chapter II, 388 were colinear between the euchromatic arms of sorghum chromosome 3 and rice chromosome 1, which indicated a level of colinearity near 85% (supplemental Table XI). The major inversion at the proximal end of sorghum chromosome 3 relative

to rice was not considered in the estimation of colinearity. The level of macrocolinearity seen in the present study is slightly higher than that previously observed by others (Klein et al., 2003; Bowers et al., 2005). This may be because in the present study, the sorghum ESTs mapped to the physical map were preselected so that they were single- or low-copy. Those sorghum ESTs that showed more than one strong hit to the rice genome by BLASTN analysis were eliminated from further consideration and therefore, the present results might be overestimating the actual level of macrocolinearity between these two homeologous chromosomes. Bowers et al. (2005) screened a sorghum physical map with sorghum ESTs to 272 rice genes spaced at ~56 kb intervals and found 72% conserved synteny among the loci. However, in their study, overgo probes were designed to the sorghum EST sequences and a hybridization-based approach was used for mapping. The design of overgo probes that will uniquely identify one loci is a difficult bioinformatic task and thus it is likely that some of the probes used by Bowers et al. (2005) may have hybridized to multiple BAC clones and/or contigs leading to the lower level of conserved synteny reported. Klein et al. (2003) used sequence scanning of BACs mapped to sorghum chromosome 3 for alignment to rice chromosome 1 and found a level of gene colinearity of ~80%. It is likely that the overall macrocolinearity between sorghum chromosome 3 and rice chromosome 1 lies somewhere between 75-85%. The release of the 8× whole sorghum genome shotgun assemblies from the JGI (Doe joint genome institute; http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html), annotation of the gene space within these assemblies and alignment to the

sorghum physical map will certainly help to resolve the true level of macrocolinearity between sorghum and rice.

Of the 456 ESTs used to construct the minimal tiling path of the euchromatic arms of sorghum chromosome 3, 68 showed a loss of colinearity with rice chromosome 1. Some of these ESTs mapped to contigs bearing the genetic markers of a linkage group other than linkage group 03 whereas others were mapped to contigs without any genetic information to confirm their correct location (Figure 4). Eight ESTs moved within chromosome 3, of which CD227587 moved within the short arm, AW563509 moved within the long arm, BE600044 moved from the long arm to the short arm and the remaining five (BG558381, CD222337, CN128558, CN139743, and BE367030) moved from the pericentromeric heterochromatic block to a euchromatic region relative to rice (supplemental Table XI). Two duplications were also detected, CD212159 mapped to a BAC located on the long arm of sorghum chromosome 3 as expected based on synteny to the orthologous gene on rice chromosome 1, however, a second copy of this gene was detected in a BAC that maps to a non-syntenic position on the short arm of chromosome 3. A similar duplication was seen with EST CN149581 where the BAC containing this gene mapped to the short arm of sorghum chromosome 3 that was colinear with the ortholog from rice and the duplicated copy mapped to a contig located on the long arm of the chromosome. Two inversions relative to rice chromosome 1 were also detected. In sorghum contig77, the orientation of the mapped ESTs was CD429994, CD212473, BG357268, BI245426, BE599043, BM322273, and BI074314, while the orientation of the orthologous genes in rice was BM322273, BI074314, BE599043, BI245426,

BG357268, CD212473, and CD429994. The data suggests that a minor inversion of BM322273 and BI074314 happened after a major inversion of the whole region. In contig26, BM325245, BG101800, and BE598591 were also inverted relative to the orientation of the orthologous genes in rice.

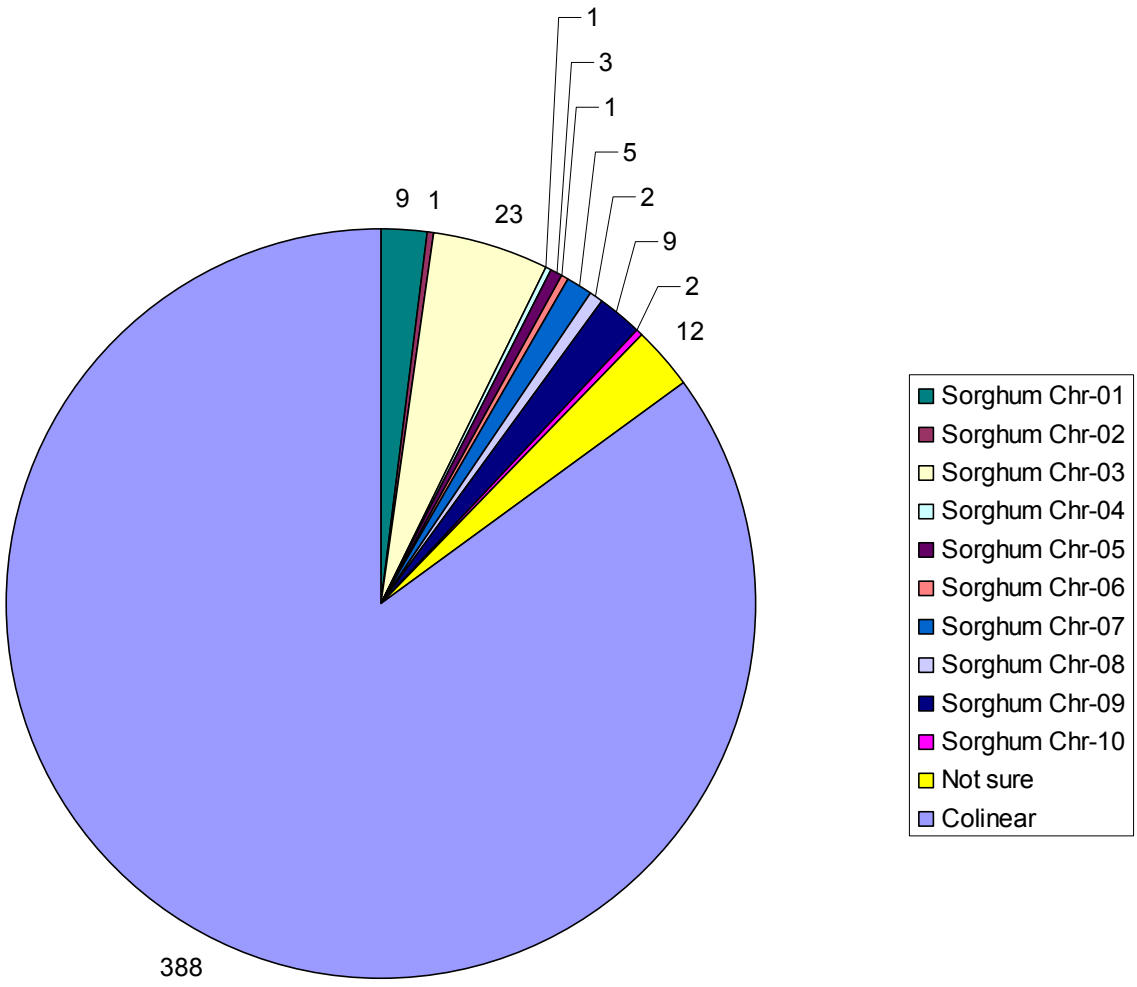

Figure 4. Distribution of colinear and non-colinear EST-STS markers. Colinearity of EST-STS markers mapped on the minimal tiling path of the euchromatic region of sorghum chromosome 3 was analyzed relative to rice chromosome 1. EST numbers of each category are shown around the pie.

Thirty-six EST-STS primers designed for the heterochromatic region produced strong and analyzable PCR signals in the BAC pools. Four contigs contained both EST-STS and AFLP markers confirming their location in the heterochromatic region and alignment to the rice chromosome 1 pseudomolecule. Additionally, two other contigs contained blocks of colinear EST-STS (3 EST-STS in one contig and 5 EST-STS in the second contig), however, these two contigs did not contain any genetic markers and therefore could not be genetically linked to the region. These results do indicate that these blocks of sorghum genes have remained in a colinear order with respect to their order in rice, however, at present it can not be determined whether these gene blocks have remained in homeologous chromosomes or not. Fifteen of the EST-STS markers appeared to have moved to other locations within the sorghum genome, 5 EST-STS markers have moved to other locations within sorghum chromosome 3 and 10 have moved to locations on other chromosomes based on the genetic markers collocated in the contigs containing these EST-STS markers. Based on the present results, the level of colinearity between the sorghum and rice euchromatic regions was ~85% whereas the colinearity within the heterochromatic regions was only ~53% (calculation includes the two contigs containing blocks of colinear EST-STS that were not genetically mapped). These results suggest that there has been a greater loss of colinearity in the gene-poor heterochromatic regions between sorghum and rice compared with the gene-dense euchromatic arms although it is likely that a larger gene sample size will be needed to confirm this. It is interesting to note that the sorghum chromosome 3 pericentromeric heterochromatic region has expanded 4.4-fold compared to the corresponding region in

rice chromosome 1, whereas the euchromatic arms of this sorghum chromosome have only expanded 1.5-fold (Kim et al., 2005a). The expansion of the heterochromatic region in sorghum relative to rice is likely due to an increase in retroelement insertion in this region in sorghum and it is possible that this activity has also resulted in gene movements and rearrangements leading to the loss of colinearity seen here.

Comparative mapping of cereal genomes using cross-hybridizing genetic markers has provided compelling evidence for a high level of conservation of gene order across regions spanning many megabases in spite of the fact that vast differences in genome size, chromosome numbers and evolutionary divergence time exist among these cereals (Bennetzen, 2000). In the present study, ~85% colinearity was identified between sorghum chromosome 3 and rice chromosome 1 by analysis of mapped EST-STS markers, although these two species have different chromosome numbers (10 for sorghum vs. 12 for rice), have ~2-fold different genome sizes and diverged from each other ~50 MYA. The initial comparative mapping studies were based on a small number of genetic markers, and thus could only resolve large chromosomal rearrangements like translocations, intra-chromosomal inversions, segmental genome duplication, and chromosomal fusion (Moore et al., 1995; Gale and Devos, 1998). In total 456 EST-STS markers across ~57.46 Mb were used in this research which provides a much higher level of resolution to estimate macrocolinearity than any previous study. Not including the major inversion at the proximal end of sorghum chromosome 3, 8 gene translocations, 2 gene duplications, 2 inversions within chromosome 3 and 45 gene

translocations to other sorghum chromosomes were detected relative to rice chromosome 1.

The present estimate of overall macrocolinearity between sorghum chromosome 3 and rice chromosome 1 may be slightly overestimated based on the methodology used, nonetheless, the high degree of gene colinearity between sorghum and rice suggests that alignment of the rice genome sequence to a high-resolution integrated genetic and physical map of sorghum will not only help illustrate the evolutionary relationship between sorghum and rice but in many instances will also accelerate the isolation and analysis of genes of agronomic importance in sorghum. In addition, this conversation of gene colinearity will be an invaluable resource in aligning the sorghum whole genome shotgun assemblies to the sorghum physical map as well as helping to guide assembly of the maize genome sequence (http://www.nsf.gov/pubs/2004/nsf04614/nsf04614.htm).

Recombination ratio of the euchromatic arms of sorghum chromosome 3

Construction of the integrated genetic and physical map of the euchromatic arms of sorghum chromosome 3 allowed the rates of recombination in these regions to be determined. The rate of recombination across the euchromatic arms varies from 0.068 Mb/cM to 0.447 Mb/cM with an average rate of 0.204 Mb/cM (Table IV). Several recombination cold and hot spots can be easily detected as shown in Figure 5.

Kim et al. (2005a) used FISH analysis to estimate that the rate of recombination of the euchromatic region of sorghum chromosome 3 and reported an average rate of 0.260 Mb/cM, while this research revealed an average rate of 0.204 Mb/cM. As described

previously, our estimate of the size of the euchromatic arms of sorghum chromosome 3 is likely an overestimate of the actual DNA content represented in the minimal tiling path. Therefore, the rate of recombination may be underestimated because the current HICF map contains BACs from two highly divergent sorghum genotypes (BTx623 and IS3620C) and to the inherent limitations associated with BAC DNA fingerprinting (Nelson et al., 2005). Although the rate of recombination estimated here and that estimated by Kim et al (2005a) are derived using different methods, they only differ by ~1.25-fold. In addition, variation in the rate of recombination across sorghum chromosome 3 is similar to that found here was also observed by Kim (J.S. Kim, personal communication).

Table IV. Recombination in the euchromatic arms of sorghum chromosome 3.
NA indicates no data were available for the pericentromeric heterochromatic block

Marker	cM	cM interval	Mb interval	Mb/cM
Xtxa3395	0.0			
Interval 1		6.6	0.94	0.142
Xtxa2105	6.6			
Interval 2		8.6	1.81	0.210
Xtxa4134	15.2			
Interval 3		7.2	1.45	0.201
Xtxa6160	22.4			
Interval 4		6.2	0.42	0.068
Xtxa2734	28.6			
Interval 5		4.7	1.04	0.221
Xtxs1092	33.3			
Interval 6		6.1	1.74	0.285
Xtxa 2096	39.4			
Interval 7		4.0	0.80	0.200
Xtxa3848	43.4			
Interval 8		5.0	1.26	0.252
Xtxa3552	49.4			
Interval 9		3.9	0.63	0.162
Xtxa267	53.3			
Pericentromeric heterochromatic block	NA	NA	NA	NA
Xtxa4054	90.6			
Interval 10		8.2	0.63	0.077

Table IV. Continued

Marker	cM	cM interval	Mb interval	Mb/cM
Xtxa2735	98.8			
Interval 11		12.1	2.45	0.202
Xtxp120	110.9			
Interval 12		14.3	2.15	0.150
Xtxa389	125.2			
Interval 13		8.1	1.58	0.195
Xtxa4019	133.3			
Interval 14		9.2	2.99	0.325
Xtxa2961	142.5			
Interval 15		8.6	3.84	0.447
Xtxa2937	151.9			
Interval 16		10.3	3.23	0.314
Xtxa2027	161.4			
Interval 17		16.0	3.42	0.214
Xtxa3891	177.4			
Interval 18		12.6	1.69	0.134
Xtxa341	190.0			
Interval 19		12.4	1.46	0.118
Xtxa4138	202.4			

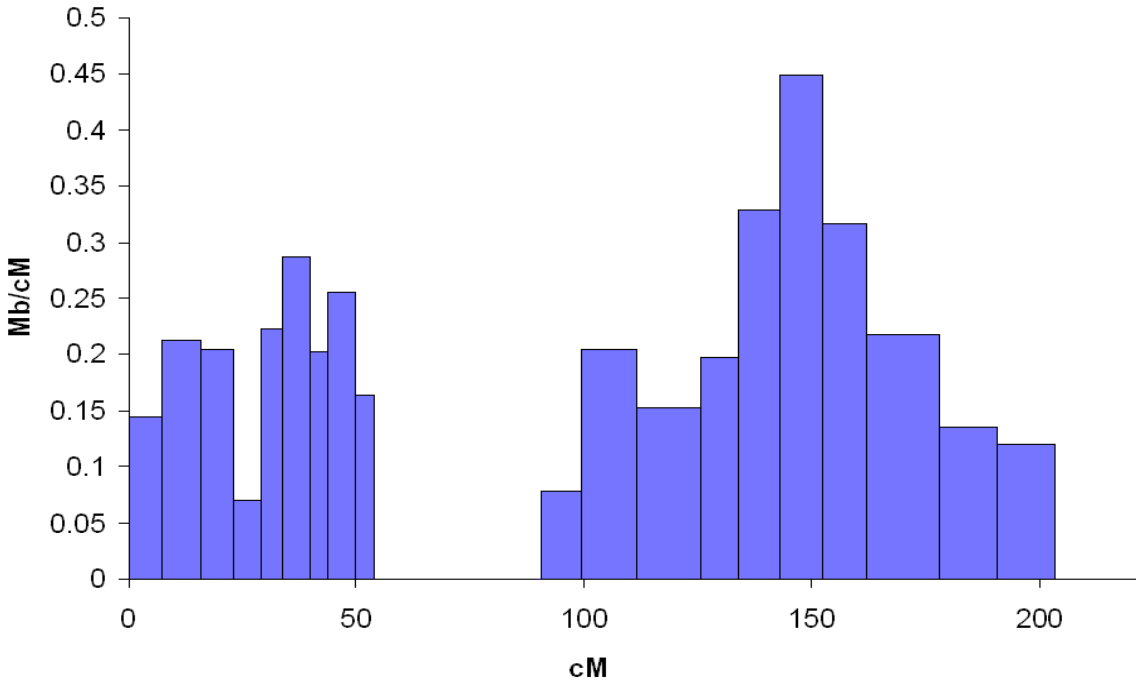

Figure 5. Recombination ratios in the euchromatic arms of sorghum chromosome 3. The pericentromeric heterochromatic block is not covered and is shown as blank in the figure.

Plant chromosomes are often divided into gene-rich and gene-poor compartments. The availability of larger and larger cereal genomic sequence datasets have confirmed the existence of the mosaic structure in which low-copy, gene-rich regions, known as 'gene islands', are interspersed among high-copy retrotransposon-rich sequences (SanMiguel et al., 1996; SanMiguel et al., 2002). It is believed that the genic regions in the cereals are associated with hot spots of recombination. The recombination ratio along the sorghum chromosome 3 euchromatic arms varied from 0.068 Mb/cM to 0.447 Mb/cM supporting the existence of a mosaic structure, which may contain low- and high-copy sequences, where the low-copy sequences tend to be gene-rich and are recombinationally active and the high-copy sequences tend to be gene-poor and recombinationally inactive. The long arm of sorghum chromosome 3 is known to contain several loci involved in abiotic stress response including *Stg1* (~140-165cM), *Stg2* (~100-120cM) and Alt_{Sb} (~174-178cM). The region around the *Stg1* locus appears to be in a region of the arm with a much lower rate of recombination compared with the *Stg2* and Alt_{Sb} loci. The reason for this lower rate of recombination around *Stg1* is unclear. However, it does suggest that fine-mapping and cloning of the underlying genes within this region might be difficult requiring the generation of very large fine mapping populations and many more markers to identify enough recombinants to narrow the locus to a small enough interval for gene identification and validation.

CHAPTER IV

MICROCOLINEARITY BETWEEN SORGHUM CHROMOSOME 3 AND RICE CHROMOSOME 1

INTRODUCTION

Over the last several years, comparative genomics in the grasses has progressed from macrocolinearity at the genetic map level toward detailed assessments of microcolinearity at the DNA sequence level. The availability of long regions of cereal genome sequence has allowed their microcolinearity to be investigated. Several studies have been carried out demonstrating that while overall macrocolinearity among the grasses is maintained, a mosaic pattern of colinearity is actually observed at the DNA sequence level (Tarchini et al., 2000; Dubcovsky et al., 2001; SanMiguel et al., 2002; Song et al., 2002; Klein et al., 2005). Numerous exceptions to colinearity including segmental duplication, gene amplification, gene loss, and gene translocation have been detected at the DNA sequence level. Regions containing fast evolving genes, such as disease resistance genes, are particularly susceptible of breaking comparative colinearity (Leister et al., 1998). The study of microcolinearity is an essential task if the information from a small, model genome species is to be used to facilitate the research of larger, more complex genome species.

To begin to explore microcolinearity between sorghum and rice in more detail two different approaches, one involving sequence skimming of overlapping BAC pools along the sorghum physical map and the other involving complete sorghum BAC sequencing were performed and the resulting sequence data compared to the orthologous regions from the rice chromosome 1 pseudomolecule.

MATERIALS AND METHODS

1.5× sequence skim and sequence analysis

The region from the long arm of sorghum chromosome 3 that aligns to rice chromosome 1 from 35 to 40.1 Mb was picked for a detailed analysis of sorghum by rice microcolinearity. The minimal tiling path of this sorghum chromosome region was constructed as described in Chapter II. Alignment to the rice chromosome 1 pseudomolecule between 35-40.1 Mb was confirmed by BAC sequence scanning (Klein et al., 2003) as well as by the EST-STS mapping detailed in Chapter II. Forty-two overlapping BAC clones from the minimal tiling path were selected and pooled into 13 individual pools each containing 4 contiguous BAC clones. The BAC clones were sent to Dr. Dick McCombie's lab at Cold Spring Harbor for the preparation of sheared libraries and sequence analysis as part of an ongoing collaboration. Each pool was sequenced to an approximate depth of 1.5×. Following collection of the sequence data at Cold Spring Harbor Laboratory, the trace and fasta files for all sequences were sent back to Texas A&M for data analysis. Sequences were assembled utilizing the PHRAP

program (http://www.phrap.org). The contig and singleton sequences generated were aligned to non-TE rice peptides (TIGR Release 3) by BLASTX (http://www.tigr.org/) analysis to determine their location and the presence of genes. Because the non-TE rice peptides have been annotated using automated annotation software many of the gene models are likely false. Therefore, a set of criteria were applied to these models to ensure that only those gene models with high confidence of actually being a gene were used in the comparative analysis to the sorghum sequences. Loci were assigned 2 points for having 3 or more significant BLASTP hits vs. GenBank, 2 points for having alignment to an EST of any grass species, 4 points for a rice cDNA alignment, and 1 point each for alignment with a sorghum or maize genome survey sequence (GSS). Loci with a total score of 2 or greater were identified as a predicted gene. The sorghum contig and singleton sequences were then subjected to BLASTX analysis against these predicted rice genes and the best hit was kept for each query if the e-value was greater than e^{-10}.

Verifying existing gaps between sorghum 1.5× pool skim sequences and high confidence rice peptides (TIGR Release 3)

Since not all high confidence rice peptides (TIGR Release 3) were aligned with the sorghum contig or singleton sequences, short gaps may exist between and inside each BAC pool. Sorghum EST or GSS sequences homeologous to these 'gap' rice peptide loci were identified and primers designed using the same strategy for EST-STS primer design as described in Chapter II. In addition, a new set of BAC pools was constructed

that covered exactly the same region of the sorghum minimal tiling path as the original pools but with different overlaps between the clones. Forty-five overlapping BAC clones from the minimal tiling path were selected and pooled into 13 individual pools each containing 2 to 5 contiguous BAC clones. The BAC clones comprising each of the original 13 pools as well as those comprising the new set of 13 pools were inoculated into 150 µl of LB broth containing 20 µg/µl chloramphenicol and grown overnight at 37C in a rotating shaker. The following day, the cultures were centrifuged to pellet the cells and BAC DNA isolated using the method of Klein et al. (1998). DNA was quantified by fluorimetry and used to screen for the presence of the EST-STS or GSS-STS markers using the same PCR conditions and settings as described in Chapter II.

Sequencing, annotation, and analysis of sorghum BACs 82G24 and 181g10

Two sorghum BAC clones, 82G24 (IS3620C genotype) and 181g10 (BTx623 genotype), from the region encompassing the BAC pools described above were selected for complete sequencing (~8× coverage) at Cold Spring Harbor Laboratory. The trace and fasta files for all sequences were sent back to Texas A&M for data analysis.

Sequences of sorghum BAC clones 82G24 and 181g10 were submitted to the Rice Genome Automated Annotation System for an initial examination (http://ricegaas.dna.affrc.go.jp/). Rice GAAS is a system that integrates programs for prediction and analysis of protein-coding gene structure (Sakata et al., 2002). Subsequent to this initial examination, sorghum sequences were submitted to the gene prediction programs FGENESH (http://www.softberry.com/berry.phtml) with the monocot training set used

for gene prediction, GENSCAN (http://genes.mit.edu/GENSCAN.html) and GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/). The criteria used to define a gene were (1) prediction as a gene from the three prediction programs listed above and (2) a sequence similarity to a putative, hypothetical, or known protein-coding sequence in the non-redundant database or in a protein database using BLASTX at an expect value greater than e^{-15}. Alternatively, a predicted open reading frame (ORF) that matched a monocot EST was defined as encoding for an unknown protein. Because EST libraries have been prepared from different species of sorghum, a homology cut-off of 97% sequence identity was required. If sorghum ESTs were hit at a homology ranging from 90% to 97%, the ORF was predicted to encode a gene if the ORF also identified other rice, maize, or sugarcane ESTs at greater than 90% homology over the same genomic region. These criteria were applied after eliminating ORFs homeologous with repetitive DNA elements such as a putative gag-pol gene, Kafirin cluster, and putative reverse transcriptase. After applying the above criteria, each predicted sorghum gene was examined for potential alignment to a specific rice chromosome. The best hit was kept for each predicted gene at an e-value of greater than e^{-15}.

Percent identity plot analysis of sorghum BAC clones 82G24 and 181G10 with their homeologous rice genomic sequences

Non-TE rice peptides (TIGR Release 3) of the regions homeologous to sorghum BAC clones 82G24 and 181g10 were obtained from http://www.tigr.org/ for gene prediction and the criteria described above to identify high confidence gene annotations

within the rice sequences were employed. PipMaker was used to obtain a percent identity plot of the alignment of genomic regions of rice and sorghum (Schwartz et al., 2000). To prevent alignment of duplicated genes or exons, percent identity plots were generated using the chaining option and by searching a single DNA strand during sequence alignment. The results of this alignment were compared to the percent identity plots obtained with both of these options turned off. FGENESH was used to define the position of exons and introns in each of the predicted genes. RepeatMasker (http://repeatmasker.org/) was used to identify repetitive elements using the Poaceae RepBase database (http://www.girinst.org). The results of this search were used to mask interspersed repeats during genome alignment with PipMaker.

RESULTS AND DISCUSSION

Microcolinearity analysis between sorghum chromosome 3 and rice chromosome 1 based on 1.5× sequence skim analysis of BAC pools

Skim sequences were assembled utilizing the PHRAP program (http://www.phrap.org) and 3681 contig and singleton sequences were generated. These sequences were aligned to high confidence non-TE rice peptides (TIGR Release 3) by BLASTX (http://www.tigr.org/) analysis to determine their location and the presence of genes. Detailed analysis results are listed in supplemental Table XII and the summarized results are listed in Table V. Overall, 729 sequences from the 13 BAC pools were identified as genic, and 472 of these were colinear relative to rice. The 257 non-colinear gene

sequences hit genes on all 12 rice chromosomes with the number varying from 8 gene hits on rice chromosome 9 to 39 on rice chromosome 5 (Table V). Further analysis of the gene sequences from the sorghum BAC pools identified 50 gene sequences that were duplicates due to the overlaps between the BAC clones selected for analysis and these were removed from the total for colinearity analysis. From the analysis of the 13 BAC pools, ~62% of the genes identified were colinear between sorghum and rice, while the colinearity of individual pools varied from ~54% in pool 3 to ~78% in pool 1 as shown in Figure 6.

Table V. Summary of BLASTX analysis of sorghum pool skim sequences to TIGR non-TE rice gene models (TIGR Release 3).

The number of non-colinear genes obtained from each pool is shown under the rice chromosome with the best e-value score. The last column indicates the total number of colinear genes obtained from each pool

	Rice Chr-01	Rice Chr-02	Rice Chr-03	Rice Chr-04	Rice Chr-05	Rice Chr-06	Rice Chr-07	Rice Chr-08	Rice Chr-09	Rice Chr-10	Rice Chr-11	Rice Chr-12	Colinear
Pool 1	0	1	1	1	3	0	0	3	0	0	1	2	43
Pool 2	3	1	0	0	7	3	1	0	3	1	0	5	50
Pool 3	1	6	4	3	2	3	1	3	1	2	1	1	33
Pool 4	1	2	4	1	3	1	2	0	0	0	0	1	25
Pool 5	3	5	1	1	3	1	0	1	0	0	2	1	28
Pool 6	3	2	3	1	3	3	2	1	0	0	1	1	36
Pool 7	1	2	0	1	6	1	0	3	1	2	2	4	53
Pool 8	7	3	5	3	3	2	2	0	1	2	1	1	39
Pool 9	2	3	5	2	3	3	1	0	1	5	0	2	41
Pool 10	3	0	4	1	5	3	0	4	0	1	2	0	30
Pool 11	1	2	1	1	1	1	0	0	1	0	1	1	31
Pool 12	1	1	3	2	0	2	1	3	0	0	4	0	39
Pool 13	2	2	1	1	0	0	0	1	0	2	1	0	24
Total	28	30	32	18	39	23	10	19	8	15	16	19	

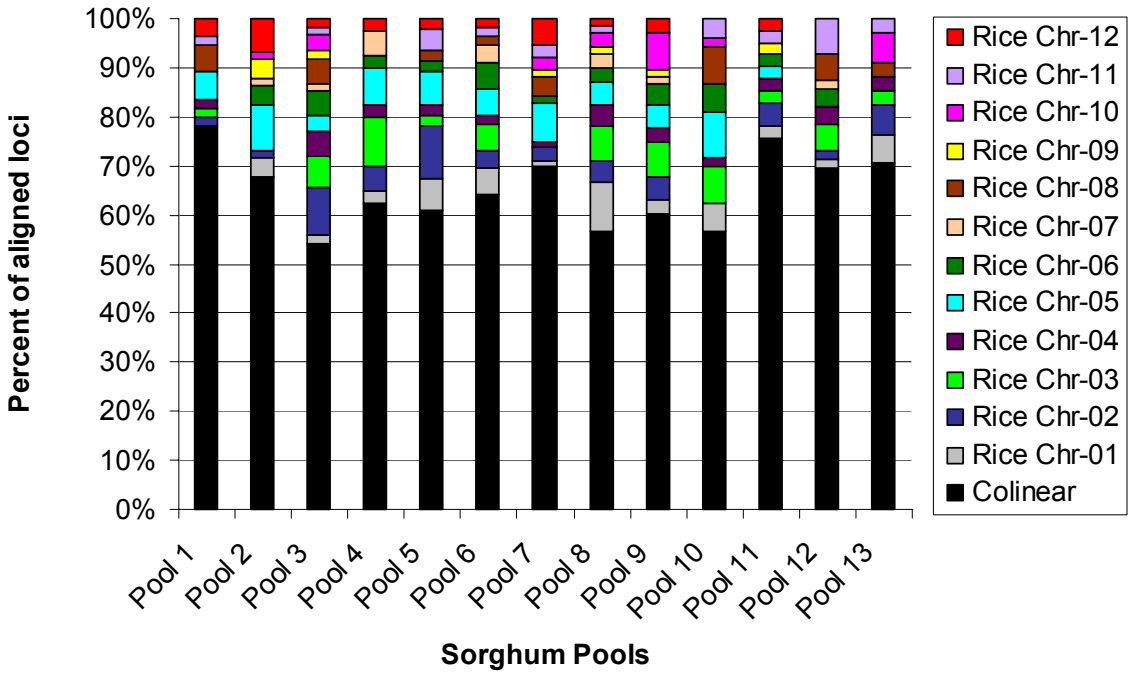

Figure 6. Sorghum BAC pool skim sequence alignment to the rice genome. Distribution of colinear and non-colinear sorghum genes relative to rice which were identified by 1.5× sequence skim of sorghum BAC pools from the minimum tiling path that aligns to rice chromosome 1 from 35 to 40.1 Mb.

The 472 colinear sorghum gene sequences present in the BAC pools were aligned to rice chromosome 1 from 35 to 40.1 Mb (Figure 7). Extensive colinearity between the minimal tiling path covered by the 13 BAC pools and rice genomic sequence was detected. Minimal overlaps between most pools were detected as expected, while a gap existed between pool 10 and pool 11, and pool 13 was almost covered by pool 12. Since not all high confidence rice peptides (TIGR Release 3) were aligned with the sorghum contig or singleton sequences, short gaps also existed inside each BAC pool. Forty-five primer sets to sorghum EST or GSS sequences homeologous to these 'gap' rice peptide

loci were designed and generated strong and analyzable signals in the BAC pools. Three primer sets designed against genes that should reside in sorghum pools 6, 9, and 10 based on the alignment to rice failed to detect any BAC clones in either the original pools used for the sequence skim or in the new pools designed to cover the region, suggesting that these genes have moved from this region relative to rice. Six pairs of primers only detected BAC clones from the new pools indicating the existence of gaps in the corresponding region in the original BAC pools (Figure 7). The original BACs for the pools were selected based on their overlap as indicated by the FPC program and it is likely that in some cases during automated contig assembly and manual merging of contigs, the overlap between BAC clones was overestimated leading to the selection, in some cases, of BACs that did not actually overlap. The remaining 36 pairs of primers detected BAC clones from the original BAC pools indicating the incomplete coverage of the sequence skim which caused genes to be missed from the analysis (Figure 7). As mentioned above, a large gap existed between sorghum BAC pools 10 and 11. This resulted because one of the BACs making up pool 10, 99b2, apparently failed to grow in the pooled culture (data not shown). Therefore, fifteen sets of primer pairs were designed to fill the gap between pool 10 and pool 11 and all of them identified this missing BAC, 99b2, which should have been in the original pool 10 (Figure 7).

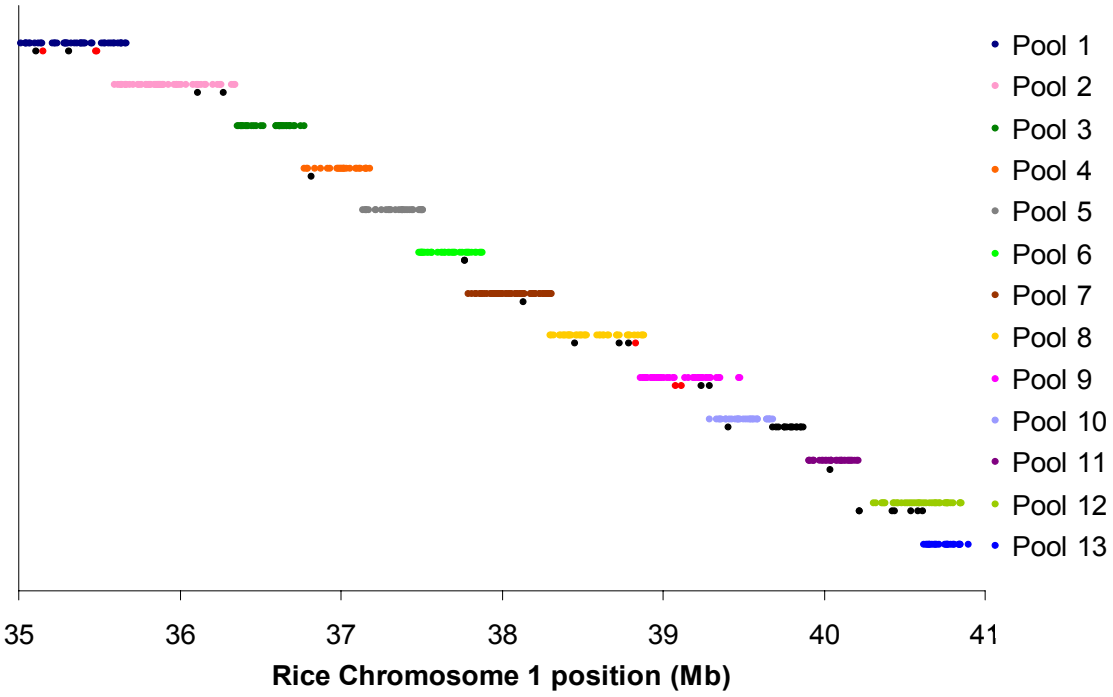

Figure 7. Sorghum pool skim sequence alignment to rice chromosome 1 from 35 to 40.1 Mb. Extensive colinearity over the targeted region between sorghum and rice was detected. The red dots indicate the real gaps in the corresponding region in the original BAC pools, while the black dots indicate the incomplete coverage of the sequence based on EST-STS mapping.

As stated previously, the 257 non-colinear gene sequences identified in the sorghum pools had BLASTX hits to all 12 rice chromosomes with the number varying from 8 genes on rice chromosome 9 to 39 genes on rice chromosome 5 (Table V). An analysis of these non-colinear gene hits to the entire TIGR rice gene dataset indicates that most of these non-colinear hits are random across the rice genome, however, there was a strong pattern of hits to one region of rice chromosome 5. Twenty-five gene sequences from pool 1 to pool 8 aligned to gene models on the rice chromosome 5 pseudomolecule from ~19.5 to 23.5 Mb (Figure 8, black dots). Further analysis of these 25 sorghum gene

sequences indicated that most of them also have a high e-value BLASTX hit to the colinear rice locus on rice chromosome 1, although the top hit is to rice chromosome 5. Upon further analysis of all of the colinear gene sequences present in the sorghum BAC pool sequence skim data, it was observed that 101 of the colinear gene sequences from pools 1 to 8 also had a high e-value hit to the rice chromosome 5 pseudomolecule from 19.5 to 23.5 Mb. These hits to rice chromosome 5 are depicted as red dots in Figure 8. These results indicate that a high level of colinearity exists between a region of sorghum chromosome 3 and rice chromosome 5.

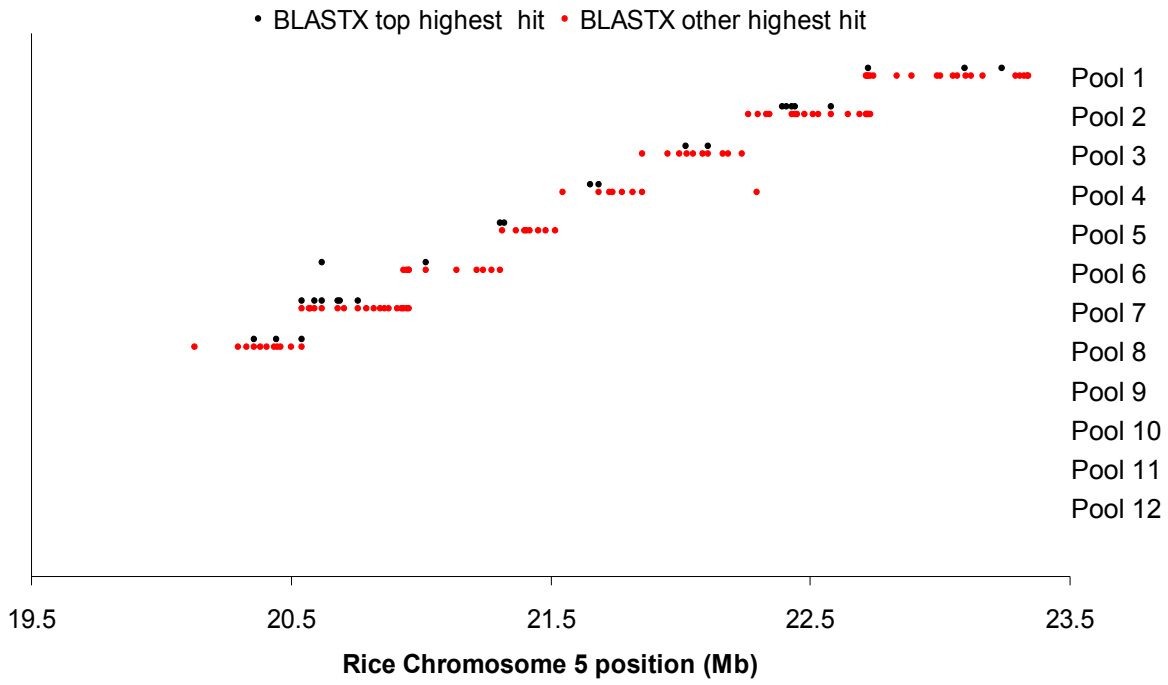

Figure 8. Cluster of sorghum pool skim sequence from pool 1 to pool 8 that align to the rice chromosome 5 pseudomolecule from 19.5 to 23.5 Mb. The top BLASTX hits to the rice chromosome 5 pseudomolecule from 19.5 to 23.5 Mb are indicated by black dots and the next highest BLASTX hits are indicated by red dots.

In recent years, the availability of sequences of long regions of the cereal genomes have allowed detailed analysis of microcolinearity to be investigated. These studies revealed that a large number of the genes in the regions analyzed were colinear but exceptions to colinearity were found as well (Avramova et al., 1996; Chen et al., 1997; Tikhonov et al., 1999 Tikhonov, 1999 #521; Dubcovsky et al., 2001; Feuillet et al., 2001; Klein et al., 2005). One large scale study of microcolinearity carried out in maize, sorghum, and two subspecies of rice found that gross macrocolinearity is maintained but that microcolinearity is incomplete among these cereals due to micro-rearrangements or small-scale genomic changes, such as gene insertions, gene deletions, gene duplications, or gene inversions (Song et al., 2002). The present results are in agreement with those of Song et al (2002). Although ~62% of the genes within the 5.1 Mb region of rice were colinear to the corresponding region of sorghum chromosome 3, there were numerous examples of a loss of colinearity including evidence for gene translocations within sorghum chromosome 3 as well as the movement of genes out of sorghum chromosome 3 relative to rice chromosome 1 (Table V and Table XII).

Extensive colinearity between the minimal tiling path covered by the 13 sorghum pools and rice genomic sequence was detected when the genic sorghum pool skim sequences were aligned to rice chromosome 1 from 35 to 40.1 Mb, although not all of the non-TE rice peptides (TIGR Release 3) from this region were aligned with sorghum contig or singleton sequences. Maximum coverage and minimal overlaps were expected when the 13 BAC pools were created. However, some small gaps actually existed in the pools based on results of screening a new set of BAC pools with primers to EST or GSS

sequences that aligned to the rice genes located within the gaps. However, many of the high confidence TIGR gene models that were not detected in the sorghum pools was likely due to incomplete coverage of the sequence skim since most of the primer sets designed to detect these genes did give positive results in BAC clones from the original BAC pools. Adding additional BAC clones into the BAC pools or closely examining the amount of overlap between neighboring BAC clones in FPC would have helped to provide better coverage of the sorghum region, however, deeper sequence analysis is also needed to get more complete coverage of the region.

A large number of non-colinear gene sequences hit rice chromosome 5 with a strong pattern, in which gene sequences present in pools 1 to 8 aligned to the rice chromosome 5 pseudomolecule from 19.5 to 23.5 Mb. Further analysis of other high e-value hits of colinear gene sequences present in these pools detected even more hits to this region of rice chromosome 5. Yu et al. (2005) also found a distinct pair of duplicated segments between 23.0 Mb to 42.3 Mb of rice chromosome 1 and 30.9 Mb to 21.6 Mb of rice chromosome 5 after analysis of improved whole-genome shotgun sequences from the genomes of *indica* and *japonica* rice. Thus it appears that the colinearity detected between sorghum chromosome 3 and rice chromosome 5 reflects the segmental homology between rice chromosome 1 and 5, which was caused by the whole genome duplication of rice predating the origination of the grasses (Yu et al., 2005).

The 1.5× sequence skim of BAC pools is an efficient and economical way for comparative genomic analysis. It provided fine scale resolution results that were not observed with the EST-STS marker strategy. Where the 1.5× sequence skim strategy

identified ~62% colinearity, the EST-STS marker strategy identified ~85% colinearity. As previously mentioned, it is likely that the EST-STS marker strategy is overestimating the true level of colinearity between sorghum and rice due to the preselection of single- or low-copy ESTs for the analysis. Although the BAC pooling method allowed us to examine the colinearity of a large contiguous region of rice and sorghum, it still can't provide the level of resolution that complete genomic sequence can. For example, the micro-rearrangements within each BAC pool, such as deletion of one member of a multi-copy gene, gene duplication, or gene inversion, can't be detected using this strategy. To examine this type of microcolinearity requires an analysis of complete genomic sequence such as that provided by 8× BAC sequencing.

Microcolinearity analysis between sorghum chromosome 3 and rice chromosome 1 based on 2 fully sequenced sorghum BAC clones, 82G24 and 181g10

Two sorghum BAC clones, 82G24 and 181g10, from the region encompassed by the 13 BAC pools were fully sequenced which permitted a detailed study of the microcolinearity of this region between rice and sorghum. The total length of BAC 82G24 is ~200 kb and the total length of BAC 181g10 is ~110 kb. The corresponding homeologous rice genomic sequences were taken from the rice chromosome 1 pseudomolecule, GenBank Accession #AP008207.

Predicted genes in sorghum BAC 82G24, 181g10 and the homeologous rice genomic sequences are shown in Table VI and Table VII, respectively. Twenty genes were predicted to be encoded in BAC 82G24, averaging one gene per 10.0 kb, while 15 genes

were predicted to be encoded in BAC 181g10, averaging one gene per 7.3 kb. Most of the predicted genes were homeologous with grass ESTs, which includes sorghum, maize, rice, and sugarcane. Eighteen and 12 genes were also predicted for the rice genomic sequences which are homeologous to sorghum BAC clones 82G24 and 181g10, respectively.

The distribution of predicted genes from sorghum BAC 82G24 and the homeologous rice genomic sequence are shown in Figure 9 and Figure 10, respectively. The distribution of predicted genes from sorghum BAC 181g10 and the homeologous rice genomic sequence are shown in Figure 11 and Figure 12, respectively. The distribution of predicted genes in BAC 82G24 was not uniform, the first ~120 kb encoded only 8 predicted genes, while the last ~80 kb encoded 12 predicted genes. In contrast, the gene distribution in BAC 181g10 is nearly uniform throughout.

Plots of the percent identity between sorghum BAC 82G24 and the homeologous rice genomic sequence are shown in the lower panels of Figure 9, while the reciprocal plots of the percent identity between the homeologous rice genomic sequence and sorghum BAC 82G24 are shown in the lower panels of Figure 10. Plots of the percent identity between sorghum BAC 181g10 and the homeologous rice genomic sequence are shown in the lower panels of Figure 11, while the reciprocal plots of the percent identity between the homeologous rice genomic sequence and sorghum BAC 181g10 are shown in the lower panels of Figure 12. Microcolinearity was largely confined to gene coding regions and sequences of exons displayed the highest percent identities. Several

Table VI. Predicted genes for sorghum BAC 82G24 and its homeologous rice genomic sequence.

Gene functions were predicted by Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/). The predicted gene transcription start and end site were predicted by FGENESH (http://www.softberry.com/berry.phtml). Blanks are left if there is no colinear gene in sorghum or rice

Sorghum BAC 82G24				Homeologous rice genomic sequence				
No.	Predicted gene function	Start	End	No.	TIGR loci	Predicted gene function	Start	End
1	Glycosyltransferase	9593	11474					
2	Triacylglycerol lipase	18366	24688	1	LOC_Os01g67420	Triacylglycerol lipase	413	7524
3	Triacylglycerol acylhydrolase	39710	40873	2	LOC_Os01g67430	Triacylglycerol acylhydrolase	19429	20736
4	Helicase	67857	76491					
5	Triacylglycerol acylhydrolase	82825	84195	3	LOC_Os01g67450	Triacylglycerol acylhydrolase	37949	39232
6	bHLH transcription factor	94631	99000	4	LOC_Os01g67480	bHLH transcription factor	53727	58456
7	Peptidase	99464	102312	5	LOC_Os01g67490	Peptidase	59086	63061
8	Ubiquitin-protein ligase	103023	106608	6	LOC_Os01g67500	Ubiquitin-protein ligase	64193	67676
9	RecA protein	120032	123718	7	LOC_Os01g67510	RecA protein	81476	86100
10	VITAMIN C DEFECTIVE 2	125837	127402	8	LOC_Os01g67520	VITAMIN C DEFECTIVE 2	87711	89252
11	4-coumarate--CoA ligase	140582	146910	9	LOC_Os01g67530	4-coumarate-CoA ligase	95263	98742
				10	LOC_Os01g67540	4-coumarate-CoA ligase	100267	103510
				11	LOC_Os01g67550	Phenol hydroxylase	105172	108647
12	Unknown protein	148372	149391					
13	Unknown protein	149655	150687					
14	Unknown protein	153684	155853	12	LOC_Os01g67560	Unknown protein	109329	111512
15	Unknown protein	157436	161433	13	LOC_Os01g67570	Unknown Protein	113076	117226
16	bHLH transcription activator	163533	164351					

Table VI. Continued

Sorghum BAC 82G24				Homeologous rice genomic sequence				
No.	Predicted gene function	Start	End	No.	TIGR loci	Predicted gene function	Start	End
				14	LOC_Os01g67580	ABC transporter	118478	125144
17	Neurolysin	166080	172856	15	LOC_Os01g67590	Neurolysin	131714	138557
18	Esterase	174130	176294	16	LOC_Os01g67600	Esterase	139538	141802
19	Unknown protein	181927	185622	17	LOC_Os01g67620	Unknown protein	147773	148471
20	Peptide transporter	197714	201577	18	LOC_Os01g67630	Peptide transporter	150076	152561

Table VII. Predicted genes for sorghum BAC 181g10 and its homeologous rice genomic sequence.

Gene functions were predicted by Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/). The predicted gene transcription start and end site were predicted by FGENESH (http://www.softberry.com/berry.phtml). Blanks are left if there is no colinear gene in sorghum or rice

Sorghum BAC 181g10				Homeologous rice genomic sequence				
No.	Predicted gene function	Start	End	No.	TIGR loci	Predicted gene function	Start	End
1	Rho-binding domain protein	5728	6442	1	LOC_Os01g68890	Rho-binding domain protein	1534	2205
2	Zinc finger protein	8434	9285	2	LOC_Os01g68900	Zinc finger protein	7629	8540
3	Unknown protein	11794	16948	3	LOC_Os01g68930	Unknown protein	27678	29516
				4	LOC_Os01g68940	Ubiquitin-like protein	33107	34783
4	Ubiquitin-like protein	22445	24754	5	LOC_Os01g68950	Ubiquitin-like protein	35905	37533
5	Nicotinic acetylcholine receptor	28942	30186					
6	Unknown protein	33766	40249					
7	CPR5 protein	42916	45402	6	LOC_Os01g68970	CPR5 Protein	40215	44987
				7	LOC_Os01g68980	Zinc finger protein	45492	46111
8	MATE efflux family protein	50016	52422	8	LOC_Os01g69010	MATE efflux family protein	64973	68243
9	Unknown protein	55149	57238					
10	Sucrose-phosphate synthase	59628	64982	9	LOC_Os01g69030	Sucrose-phosphate synthase	71328	78158
11	Ring finger protein	70427	72316	10	LOC_Os01g69040	RING finger protein	88698	92037
12	Unknown protein	81532	83165	11	LOC_Os01g69050	Unknown protein	98862	103401
13	Unknown protein	85100	85984					
14	Aminotransferase-like protein	95091	98357					
15	Hydrolase-like protein	105387	109078	12	LOC_Os01g69060	Hydrolase-like protein	104390	108060

exceptions were also detected where the intergenic regions did show a detectable level of microcolinearity, while there were no predicted protein-coding sequences residing in these regions.

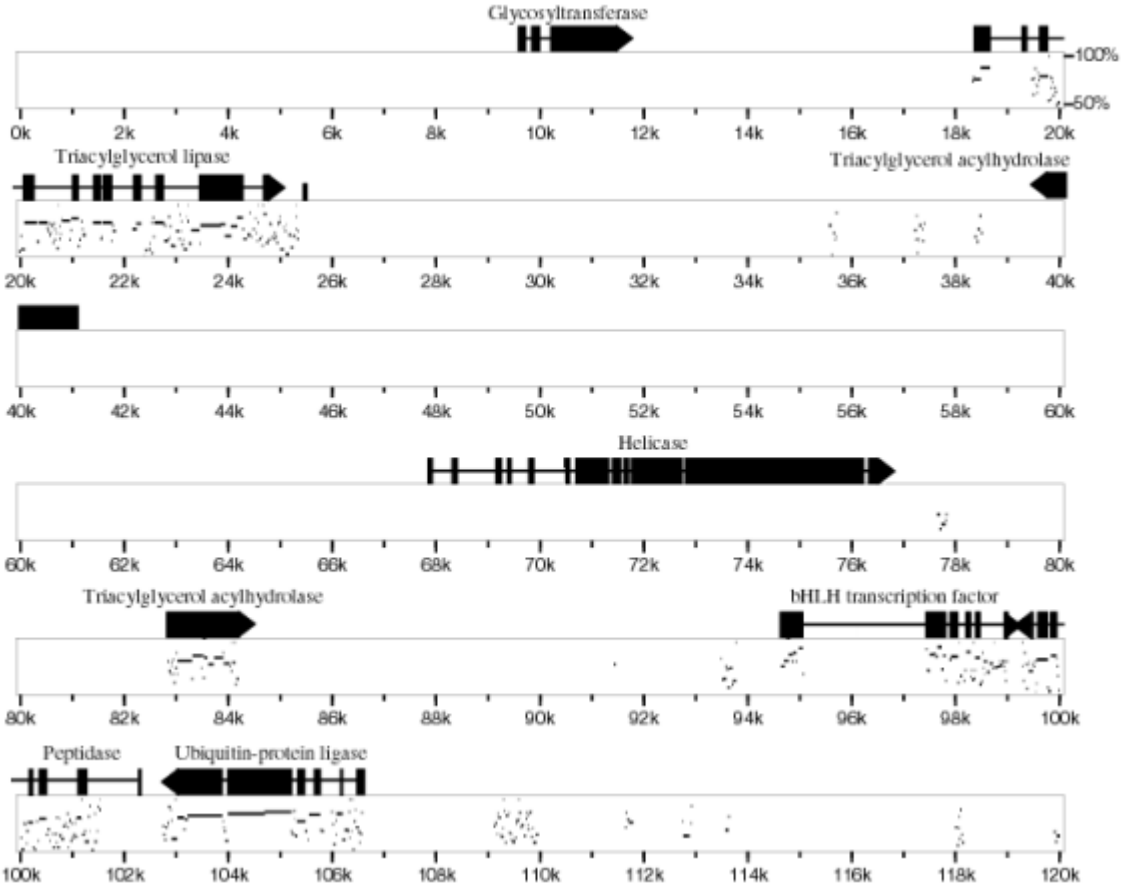

Figure 9. Gene map of sorghum BAC 82G24 and percent identity plot with a 154 kb homeologous segment of the rice chromosome 1 pseudomolecule. Predicted genes are listed in Table VI. Transcription orientation of predicted genes is indicated by arrowheads. Exons are indicated as black rectangles and introns by thin lines between them. Percent identity plots showing the alignment of sorghum and rice sequences is shown in the light gray boxes.

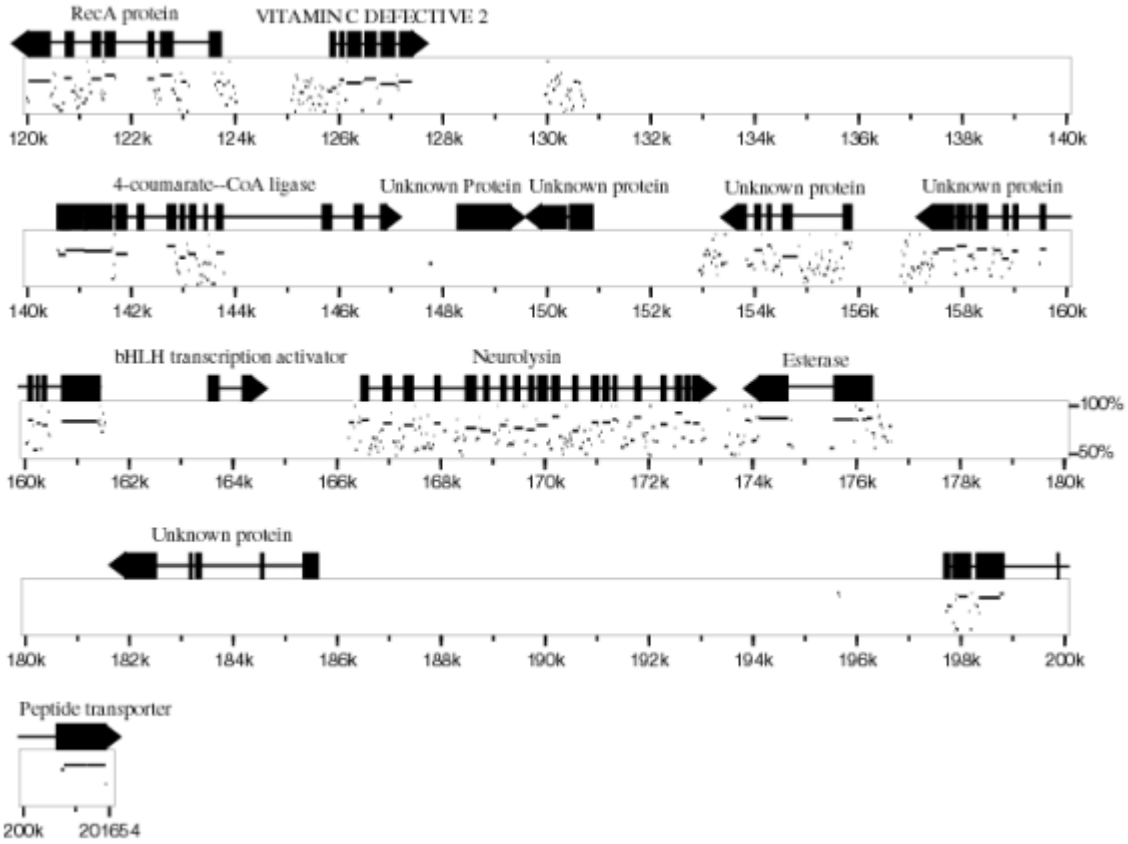

Figure 9. Continued

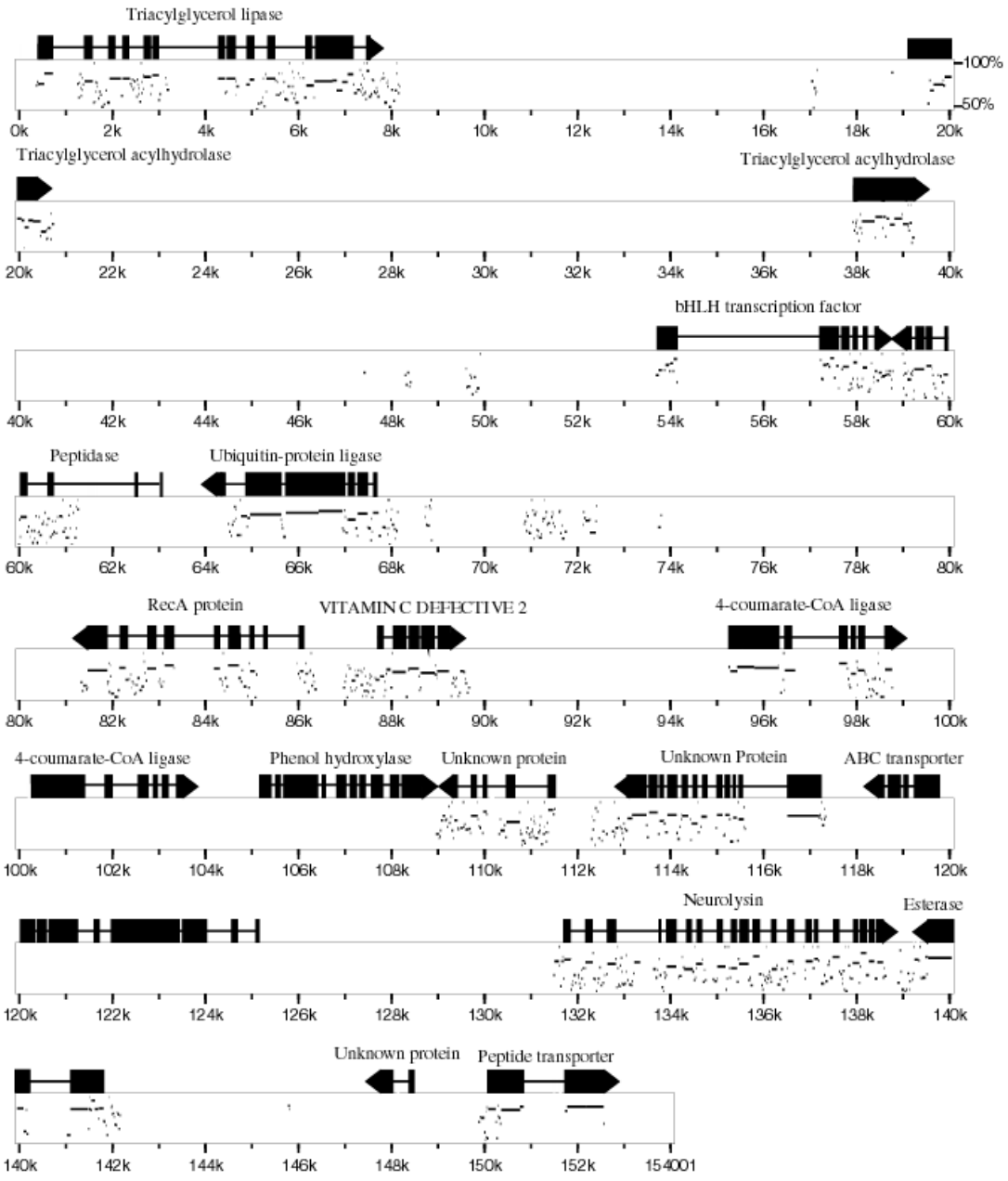

Figure 10. Gene map of a 154 kb segment of the rice chromosome 1 pseudomolecule and percent identity plot with the homeologous sorghum BAC 82G24. Predicted genes are listed in Table VI. Transcription orientation of the predicted genes is indicated by arrowheads. Exons are indicated as black rectangles and introns by thin lines between them. Percent identity plots showing the alignment of rice and sorghum sequences is shown in the light gray boxes.

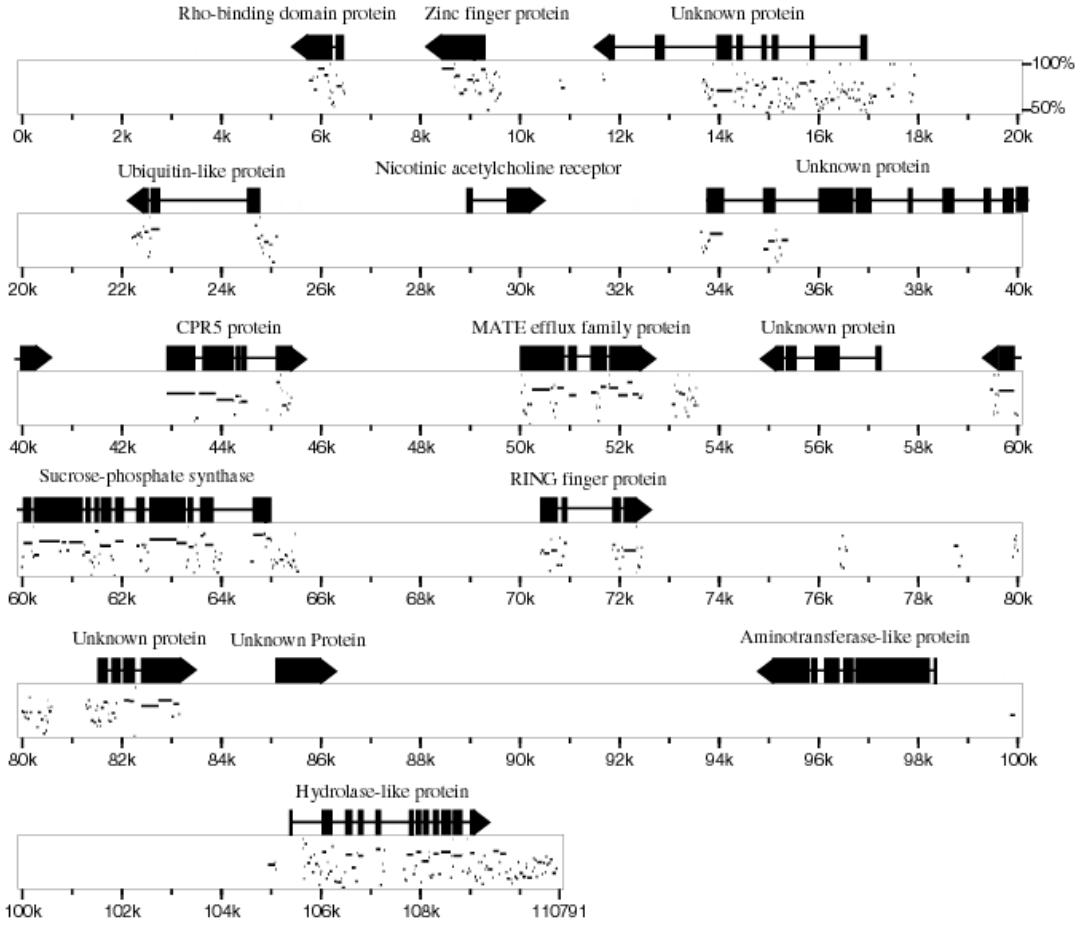

Figure 11. Gene map of sorghum BAC 181g10 and percent identity plot with a 110 kb homeologous segment of the rice chromosome 1 pseudomolecule. Predicted genes are listed in Table VII. Transcription orientation of predicted genes is indicated by arrowheads. Exons are indicated as black rectangles and introns by thin lines between them. Percent identity plots showing the alignment of sorghum and rice sequences is shown in the light gray boxes.

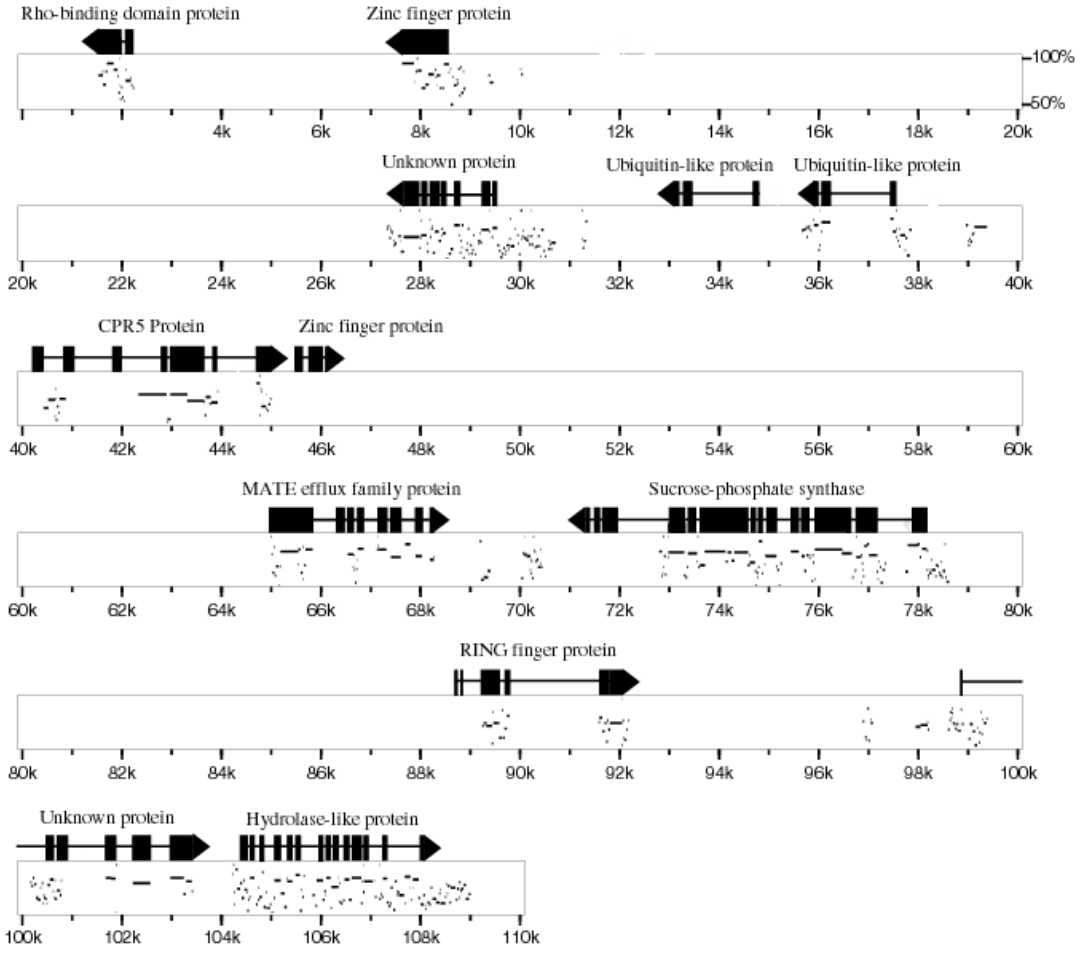

Figure 12. Gene map of a 110 kb segment of the rice chromosome 1 pseudomolecule and percent identity plot with homeologous sorghum BAC 181g10. Predicted genes are listed in Table VII. Transcription orientation of predicted genes is indicated by arrowheads. Exons are indicated as black rectangles and introns by thin lines between them. Percent identity plots showing the alignment of rice and sorghum sequences is shown in the light gray boxes.

Duplication of two genes in rice relative to sorghum was detected in the sequenced regions. When the homeologous rice sequence is compared to the sequence of sorghum BAC 82G24, a predicted 4-coumarate-CoA ligase is duplicated (Figure 10). Similarly, a duplication of the predicted ubiquitin-like protein has occurred in rice relative to sorghum as seen from the annotation of sorghum BAC 181g10 (Figure 12). These 2 genes likely underwent duplication after the divergence of the rice and sorghum ancestors since the homologues were not found in the same relative position in sorghum.

Changes in gene orientation were also detected when comparing the rice and sorghum sequences. The first predicted triacylglycerol acylhydrolase (Figure 9) may reside in a colinear position in sorghum and rice, but the transcription orientation appears opposite in these two species. No similarity was detected for this gene between sorghum and rice because the percent identity plots were generated using the chaining option and by searching a single DNA strand during sequence alignment to prevent alignment of duplicated genes or exons. However, microcolinearity was detected when both of these options were turned off.

When sorghum BAC 82G24 was compared with the homeologous rice genomic sequence, 15 out of the 20 predicted genes were colinear corresponding to 75% colinearity (Figure 9). The non-colinear predicted glycosyltransferase and predicted helicase have homologs located in other rice chromosomes which may indicate their translocation to new chromosomal locations after the divergence of sorghum and rice. In contrast, the 3 additional non-colinear predicted genes in BAC 82G24 have no homologs within the rice genome and may indicate the loss of these genes in rice

compared to sorghum. When the homeologous rice genomic sequence was compared with sorghum BAC 82G24, 15 out of the 18 predicted genes in the rice sequence were also predicted in the colinear order in BAC 82G24 indicating a level of colinearity of ~83% (Figure 10). If the two analyses are taken together, 15 out of 23 predicted genes were colinear between the sorghum and rice sequences giving an overall level of colinearity of ~65%.

When sorghum BAC 181g10 was compared with the homeologous rice genomic sequence, 10 out of the 15 predicted genes (~67%) were colinear. The 5 non-colinear predicted genes have no homolog anywhere in the rice genome and again may indicate the loss of these genes in rice relative to sorghum. When the homeologous rice genomic sequence was compared with sorghum BAC 181g10, 10 out of the 12 predicted genes were colinear indicating ~83% colinearity (Figure 12). If the two analyses are taken together, 10 out of the 17 predicted genes (~59%) were colinear.

The average gene density of sorghum euchromatic regions is predicted to be one gene for every 12.3 kb, assuming there are a similar number of genes in colinear regions of sorghum and rice (Kim et al., 2005a). Sorghum BAC clones 82G24 and 181g10 were selected from the euchromatic region of the long arm of chromosome 3 and the gene density in these two BACs is one gene per 10.0 kb and one gene per 7.3 kb, respectively. The gene density reported here is higher than the average gene density reported by Kim et al. (2005a) and likely reflects the fact that these BAC clones are from a region of the genome with higher gene content. From the analysis of recombination along the sorghum chromosome 3 tiling path (Chapter III, Table IV and Figure 5), it is clear that a

mosaic organization of genes interspersed with repetitive DNA exists and that both regions of high and low gene-density are present. This is exactly what has been observed on other sorghum chromosomes. For example, Klein et al. (2005) sequenced a 0.53 Mb region of sorghum chromosome 8 surrounding the *Rf1* locus and found significant differences in gene density even within this small region. The *Rf1* region could be divided into a gene-poor region with a gene density of 1gene/ ~45 kb followed by a gene-rich region with a gene density approaching 1 gene/~10 kb.

Small-scale gene rearrangements including gene duplication, gene loss, gene translocation, and gene inversion have been detected in the present study which is in agreement with previous findings (Tarchini et al., 2000; Dubcovsky et al., 2001; SanMiguel et al., 2002; Song et al., 2002; Klein et al., 2005). The duplicated genes may degenerate into pseudogenes or evolve distinct functions (Lynch and Conery, 2000), as seen in clusters of disease resistance (Meyers et al., 1998; Feuillet and Keller, 1999). At present, the mechanism of single-gene translocation is unknown. One putative mechanism is an ancient gene duplication in the common ancestor followed by the loss of one gene copy in the first modern species and the loss of the other copy in the second species, which functionally associated the gene amplification and gene translocation together (Song et al., 2002). In the future, detailed comparative sequence analysis of the orthologous region of maize which diverged from sorghum ~16 MYA may help to address this question.

When sorghum BAC 82G24 was compared with the homeologous rice genomic sequence, 8 out of the 23 predicted genes were not colinear, of which 5 were present in

the sorghum sequence but not in the homeologous rice sequence whereas 3 were present in the rice sequence but not in sorghum (Figures 9 and 10). When sorghum BAC 181g10 was compared with the homeologous rice genomic sequence, 7 out of the 17 predicted genes were not colinear, of which 5 were present in sorghum BAC 181g10 but not in the homeologous rice sequence and 2 were present in the rice sequence but not in BAC 181g10 (Figure 11 and 12). Sorghum BAC clones 82G24 and 181g10 and their orthologous rice genomic sequences were compared by PipMaker to obtain the percent identity plots which provided insight into the modes of local sequence evolution that have occurred over the past ~50 million years. Regions of microcolinearity reflected gene-coding sequences and the regions immediately flanking these genes, whereas gaps in colinearity reflected missing genic sequences and intergenic spaces. Microcolinearity between sorghum and rice was also detected in several intergenic regions that bore no predicted protein-coding sequences. These short conserved sequences outside of the coding exons presumably may be part of important regulatory regions (Hardison, 2000) or related with certain unknown functional significance (Kaplinsky et al., 2002).

It has been postulated that rice can act as a surrogate model for the cloning of genes in species with larger, more complex genomes such as sorghum and maize. Our results, like those of others (Tarchini et al., 2000; Dubcovsky et al., 2001; SanMiguel et al., 2002; Song et al., 2002; Klein et al., 2005), show that even in regions of high microcolinearity, exceptions exist due to small-scale gene rearrangements including gene loss and gene translocation. For instance, Klein et al (2005) used a map-based approach to clone the sorghum *Rf1* gene and compared the locus to the syntenic region

of rice. In their analysis, they were able to observe significant colinearity between sorghum and rice, however, an ortholog of the sorghum *Rf1* gene (PPR13) was not present in the colinear region of rice. Thus cloning of this gene from sorghum would not have been possible through a candidate gene approach using rice as a guide. Hence, it will be an important task to develop the genomic resources for those related grass species to complement the resources already developed for rice. The study of microcolinearity between these species will be essential if the information from one species is used to facilitate the research of another species.

CHAPTER V

EXAMINATION OF THE EXPRESSION PATTERNS OF COLINEAR AND NON-COLINEAR GENES

INTRODUCTION

Macrocolinearity and microcolinearity analysis between sorghum and rice have detected numerous gene movements. It is possible that gene movement could be associated with a change in the transcriptional regulation of that gene in one species compared to another. For example, if a gene is relocated into a region with novel genetic or epigenetic features, it's expression pattern might be altered, as has often been observed when a transposable element inserts near a gene thereby altering the regulation of its expression (Barkan and Martienssen, 1991). Some translocated genes may even move to different chromosomes and the entirely new chromosomal environment may cause a new expression pattern. This hypothesis is supported by an analysis of the *zein* region of maize, where the recently amplified new gene copies showed a change in their transcriptional regulation by a different transcription factor (Song et al., 2001). There are also examples where the movement of a gene has not resulted in a change in its expression pattern. Examination of the alcohol dehydrogenase1 (*adh1*)-orthologous regions of rice and maize suggested that *adh1* was transposed as a single gene to a new location in the *Andropogoneae*, but no change in the basic exon/intron organization or

alteration in the tissue specificity of *adh1* expression resulted from this transposition (Tarchini et al., 2000; Bennetzen and Ramakrishna, 2002).

To examine whether the loss of gene colinearity between sorghum and rice might result in a change in the expression pattern of the gene, analysis of a select number of colinear and non-colinear genes from sorghum and rice was performed using qRT-PCR because this technique has excellent sensitivity, specificity and dynamic range. In this analysis, roots and shoots from both control and ABA-treated rice and sorghum seedlings were examined, as changes in gene expression might be tissue- and/or treatment-specific. Treatment of seedlings with ABA was chosen for an initial analysis since this hormone is induced by various abiotic stresses such as salt, cold, drought and wounding leading to the induction of hundreds of different genes (Zhu, 2002; Buchanan et al., 2005). ABA is also indispensable in hormonal regulation of seed germination and further developmental processes (Cheng et al., 2002; Finkelstein et al., 2002; Barrero et al., 2005). Additionally, the analysis included two sorghum as well as two rice genotypes to address whether genotypic differences might also be present. Colinear genes were selected from the two fully sequenced sorghum BAC clones, 82G24 and 181g10, and the non-colinear genes were selected from sorghum BAC pool skim sequences.

MATERIALS AND METHODS

Plant growth and ABA treatment

Two sorghum genotypes, BTx623 and IS3620C, and two rice genotypes, Nipponbare and Lemont, were utilized in these experiments. Seeds were placed in glass beakers containing distilled water, aerated with a bubbler for 24 h and then sterilized and germinated for 3 days on germination paper (Anchor Paper, St. Paul, MN). Following germination, the seedlings were transferred to custom hydroponic tanks holding 40 plants each. Seedlings were grown hydroponically under constant aeration in 0.5× Hoagland's nutrient solution in a growth chamber at 31°C day, 22°C night temperature. Day length was 12 h, humidity was constant at 50%, and nutrient solution was replenished on day 6 and day 12. ABA ((±)-cis, trans-abscisic acid, Sigma, St. Louis, MO) was used for treatment by dissolving the appropriate volume of stock solution into the nutrient solution to obtain a final concentration of 125 μM. Sorghum was grown for 9 days and rice for14 days before the treatment to keep them at the same developmental stage. For each genotype and treatment, there were three biological replicates. These replicates consisted of independent, spatially separated hydroponic vessels in which the seedlings were grown at the same time within the same growth chamber. Twenty-seven h following ABA-treatment, tissues were harvested from both control and treated samples from a pool of at least 10 plants per replicate. Shoots and roots were quickly divided at the residual seed coat, flash frozen in liquid Nitrogen and stored at -80 °C.

The timing of germination, seed transfer, feeding, treatments and collection were kept constant among experiments to reduce the impact of circadian variation.

Quantitative RT-PCR analysis

RNA was extracted using a Trizol-based RNA extraction method (Molecular Research Center, Cincinnati, OH). RNAs were converted to cDNA template for qRT-PCR using random hexamer primers and Multiscribe reverse transcriptase (Applied Biosystems, Foster City, CA). qRT-PCR was performed in duplicate 10 μL reactions using Sybr Green mastermix (Applied Biosystems, Foster City, CA) for the sample reactions and ribosomal control reactions. Non-template control reactions using un-transcribed RNA controls confirmed that no interfering products derived from genomic DNA were present. Primers for amplifying genes of interest were designed using PrimerQuest^{TM} provided by www.idtdna.com. All primers were initially tested for primer efficiency. Serial dilutions of first strand cDNA to a final concentration of 10 ng/μL, 5 ng/μL, 2.5, ng/μL 1.25, and 0.625 ng/μL RNA equivalent were made and used as templates for RT-PCR analysis. A standard curve was plotted using Log (starting cDNA concentration) as the X-axis and threshold cycle as the Y-axis. Slope and R^2 were calculated and primer efficiency was calculated as $(10^{-1/slope}-1)\times 100\%$. The primers used for final analysis were required to have primer efficiency greater than 75% and differ no more than 5% between rice and sorghum. Amplification specificity was determined by dissociation curve analysis. Mean induction folds were calculated as $2^{(\Delta\Delta CT)}$, and SD range of replicate reactions was calculated by: upper error bar = $2^{(\Delta\Delta CT+s)}$,

lower error bar = $2^{(\Delta\Delta CT - s)}$, where: $\Delta\Delta CT = (\Delta CT_{control\ cDNA}) - (\Delta CT_{treatment\ cDNA})$, $\Delta CT = (\text{mean CT cDNA}_{test\ primers}) - (\text{mean CT cDNA}_{ribosomal\ primers})$, $S = \sqrt{[(\text{sd of } CT_{test\ primers}^2) + (\text{sd of } CT_{ribosomal\ primers}^2)]}$.

RESULTS AND DISCUSSION

Quantitative RT-PCR analysis of the selected colinear and non-colinear genes

RT-PCR was used to study the gene expression patterns of colinear and non-colinear genes from sorghum and rice. Representative colinear genes were selected from the two fully sequenced sorghum BAC clones, 181g10 and 82G24 (Table VI and VII), while non-colinear genes were selected from the sequences obtained from the 13 BAC pools used for the 1.5× sequence skim analysis (supplemental Table XII). The expression level of these genes under control and ABA treatment are shown in Table VIII and analyzed results of fold expression change after ABA treatment are shown in Table IX together with the ratio of basal gene expression between the shoot and root tissue in control, untreated plants. Twelve genes that showed colinearity between sorghum and rice were examined, whereas only 5 non-colinear genes were examined (Table VIII and IX).

Table VIII. Expression level of selected sorghum and rice colinear and non-colinear genes under control and ABA treatment.

Predicted genes from sorghum BAC clones, 181g10 and 82G24, are named the same as Table VI and Table VII. Predicted genes from the sequences obtained from the 13 BAC pools used for the 1.5× sequence skim analysis are named using the pool and contig number. No RT-PCR results were available for some primer sets if they failed to meet the primer efficiency criteria and are indicated by NA

	Control								ABA treatment							
	Shoots[a]				Roots[a]				Shoots[a]				Roots[a]			
Predicted gene	BTx	IS	Ni	Le	BTx	IS	Ni	Le	BTx	IS	Ni	Le	BTx	IS	Ni	Le
181G10_1	20.3	18.9	19.8	19.5	19.3	18.8	19.8	19.5	18.9	18.3	20.5	19.8	20.7	20.6	21.5	21.3
181G10_2	20.5	20.9	21.6	21.4	22.3	22.2	21.6	21.3	21.5	21.3	22.7	21.8	24.6	24.7	22.5	22.2
181G10_3	22.2	22.8	19.6	18.9	22.4	22.8	19.8	18.7	22.5	22.8	20.2	18.8	23.0	23.5	21.9	20.0
181G10_7	19.3	19.5	21.1	20.0	18.9	19.7	21.4	20.2	19.2	19.4	20.5	19.3	18.6	19.7	21.0	20.1
181G10_10	16.5	17.2	NA	18.5	23.3	23.3	NA	23.9	16.9	18.1	NA	18.2	18.9	21.5	NA	20.8
181G10_12	20.7	NA	17.6	17.5	21.5	NA	19.6	19.6	20.8	NA	17.8	17.5	21.0	NA	18.4	18.6
82G24_2	25.8	NA	23.3	NA	26.1	NA	23.7	NA	25.6	NA	23.2	NA	26.3	NA	22.2	NA
82G24_7	NA	19.5	21.9	20.1	NA	20.5	25.5	22.9	NA	19.1	23.2	21.0	NA	21.4	27.3	24.1
82G24_8	20.00	20.6	21.7	19.8	18.5	18.7	19.9	18.5	18.9	20.1	21.9	19.8	19.7	20.1	22.2	19.8
82G24_15	18.1	18.7	19.2	18.0	24.7	25.9	23.0	21.3	18.8	20.0	19.7	18.2	24.9	25.5	22.7	21.5
82G24_18	20.5	20.9	NA	19.6	20.2	20.7	NA	19.4	21.3	21.1	NA	19.2	20.4	20.8	NA	20.2
82G24_20	19.9	21.3	24.7	25.0	22.0	25.7	24.4	24.4	22.7	22.3	24.9	23.6	26.7	30.7	23.2	21.9
Pool9_ctg186	19.0	20.5	20.0	20.4	19.7	20.7	20.3	19.6	18.2	19.3	19.7	19.3	18.0	19.7	20.0	19.4
Pool10_ctg52	21.5	21.6	20.9	20.0	21.0	21.0	21.8	21.0	20.2	20.8	19.7	19.3	20.8	20.8	19.2	19.3
Pool10_ctg173	NA	22.2	20.6	NA	NA	22.8	20.6	NA	NA	22.3	20.0	NA	NA	23.0	19.8	NA
Pool11_ctg26	22.2	24.0	26.5	24.4	22.1	23.5	27.3	24.7	22.6	23.7	26.3	24.0	23.3	24.7	26.9	24.4
Pool11_ctg75	20.8	20.6	20.3	20.1	22.3	22.1	20.6	20.7	20.9	21.2	19.5	19.4	21.9	22.1	20.2	20.3

[a] BTx= BTx623, IS= IS3620C, Ni= Nipponbare, Le= Lemont

Table IX. Fold of expression change for selected sorghum and rice colinear and non-colinear genes under ABA treatment.

Predicted genes from sorghum BAC clones, 181g10 and 82G24, are named the same as Table VI and Table VII. Predicted genes from the sequences obtained from the 13 BAC pools used for the 1.5× sequence skim analysis are named using the pool and contig number. Gene functions were predicted by Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/). Positive values indicate the fold of gene induction after ABA treatment, while the fold of repression is represented as a negative value. No RT-PCR results were available for some primer sets if they failed to meet the primer efficiency criteria and are indicated by NA

		S/R [a b]				Shoots[a]				Roots[a]				
Predicted gene	Colinear	BTx	IS	Ni	Le	BTx	IS	Ni	Le	BTx	IS	Ni	Le	Predicted function
181G10_1	Yes	2.0	-1.1	1.0	1.0	2.6	1.5	-1.7	-1.2	-2.7	-3.3	-3.2	-3.5	Rho-binding domain protein
181G10_2	Yes	-3.5	2.5	1.0	-1.1	-2.0	-1.3	-2.1	-1.3	-4.8	-5.8	-1.9	-1.9	Zinc finger protein
181G10_3	Yes	-1.1	1.0	1.1	-1.1	-1.2	1.0	-1.5	1.0	-1.5	-1.6	-4.3	-2.4	Unknown protein
181G10_7	Yes	1.3	1.1	1.2	1.1	1.1	-1.1	1.5	1.6	1.2	1.0	1.3	1.1	CPR5 protein
181G10_10	Yes	111.4	68.6	NA	42.2	-1.3	-1.9	NA	1.2	21.3	3.7	NA	8.5	Sucrose-phosphate synthase
181G10_12	Yes	1.7	NA	4.0	4.3	-1.1	NA	-1.1	1.0	1.4	NA	2.3	2.1	Unknown protein
82G24_2	Yes	1.2	NA	1.3	NA	1.2	NA	1.0	NA	-1.1	NA	2.7	NA	Triacylglycerol lipase
82G24_7	Yes	NA	2.0	12.1	7.0	NA	1.3	-2.4	-1.9	NA	-1.7	-3.4	-2.4	Peptidase
82G24_8	Yes	-2.8	-3.7	-3.5	-2.5	2.0	1.4	-1.2	1.0	-2.3	-2.6	-4.9	-2.5	Ubiquitin-protein ligase
82G24_15	Yes	97.0	147.0	13.9	9.8	-1.7	-2.5	-1.4	-1.1	-1.1	1.3	1.2	-1.1	Unknown protein
82G24_18	Yes	-1.2	-1.1	NA	-1.1	-1.7	-1.1	NA	1.3	-1.2	1.0	NA	-1.9	Esterase
82G24_20	Yes	4.3	21.1	-1.2	-1.5	-6.9	-2.1	-1.1	2.6	-27.2	-32.1	2.4	5.6	Peptide transporter
Pool9_ctg186	No	1.6	1.1	1.2	-1.7	1.8	2.2	1.2	2.1	3.2	2.1	1.3	1.2	Glycosyltransferase
Pool10_ctg52	No	-1.4	-1.5	1.9	2.0	2.5	1.8	2.2	1.7	1.1	1.1	5.7	3.4	2'-hydroxyisoflavone reductase
Pool10_ctg173	No	NA	1.5	1.0		NA	-1.1	1.4	NA	NA	-1.1	1.8	NA	Unknown protein
Pool11_ctg26	No	-1.1	-1.4	1.7	1.2	-1.3	1.2	1.1	1.4	-2.3	-2.3	1.3	1.3	Unknown protein
Pool11_ctg75	No	2.8	2.8	1.2	1.5	-1.1	-1.5	1.7	1.7	1.3	1.0	1.4	1.3	Unknown protein

[a] BTx= BTx623, IS= IS3620C, Ni= Nipponbare, Le= Lemont

[b] S/R is the ratio of basal gene expression between the shoot and root tissue in control, untreated plants.

Expression differences were detected, some were species specific (rice vs. sorghum) and some were tissue specific (shoot vs. root), and some showed a mixed pattern of changes. For example, the expression of predicted gene 82G24_20 was species specific (rice vs. sorghum), 2 sorghum varieties showed repressed expression after ABA treatment while 2 rice varieties showed unchanged (Nipponbare shoot) or induced expression (Figure 13). In another example, a tissue specific (shoot vs. root) expression change was detected. The expression of predicted gene 181g10_10 was almost unchanged in shoot after ABA treatment for the three varieties (BTx623, IS3620C and Lemont) examined, while an apparent induction was detected for the roots of these three varieties (Figure 14). These 2 predicted genes are colinear between sorghum and rice and other colinear genes showed similar or mixed patterns of expression changes after ABA treatment. No apparent patterns could be detected from the 5 non-colinear predicted genes because they all showed mixed patterns of expression change (Table IX). Based on the current RT-PCR analysis, none of the expression pattern changes observed could be directly linked to the loss of colinearity.

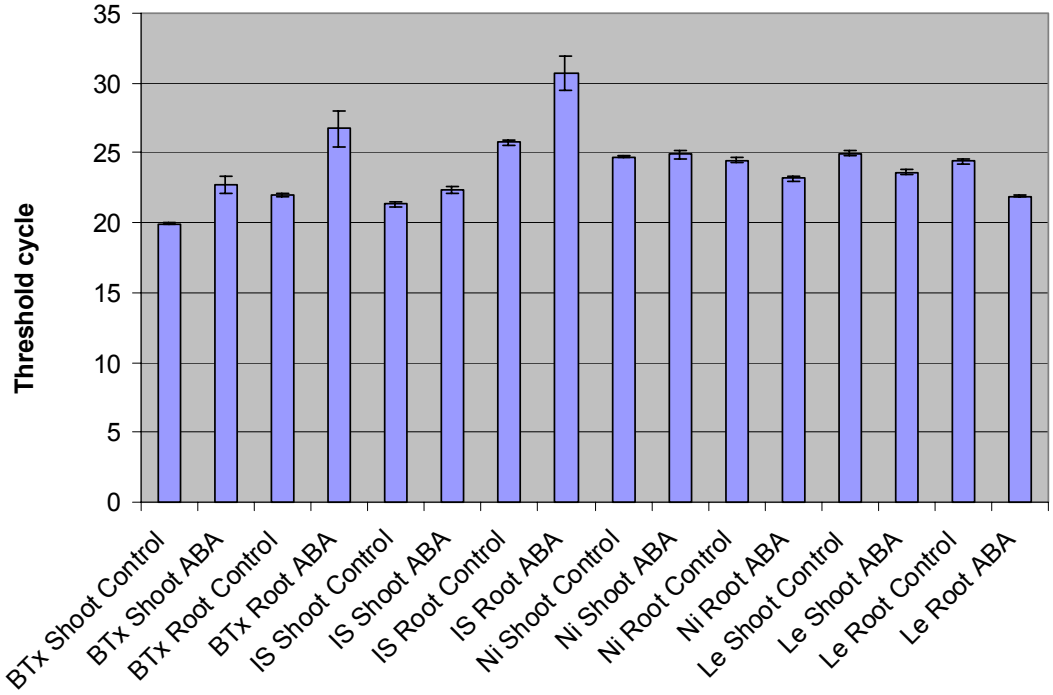

Figure 13. Expression level of predicted gene 82G24_20 under control and ABA treatment. SD range is calculated by upper error bar and lower error bar on top of each column. BTx= BTx623, IS= IS3620C, Ni= Nipponbare, Le= Lemont

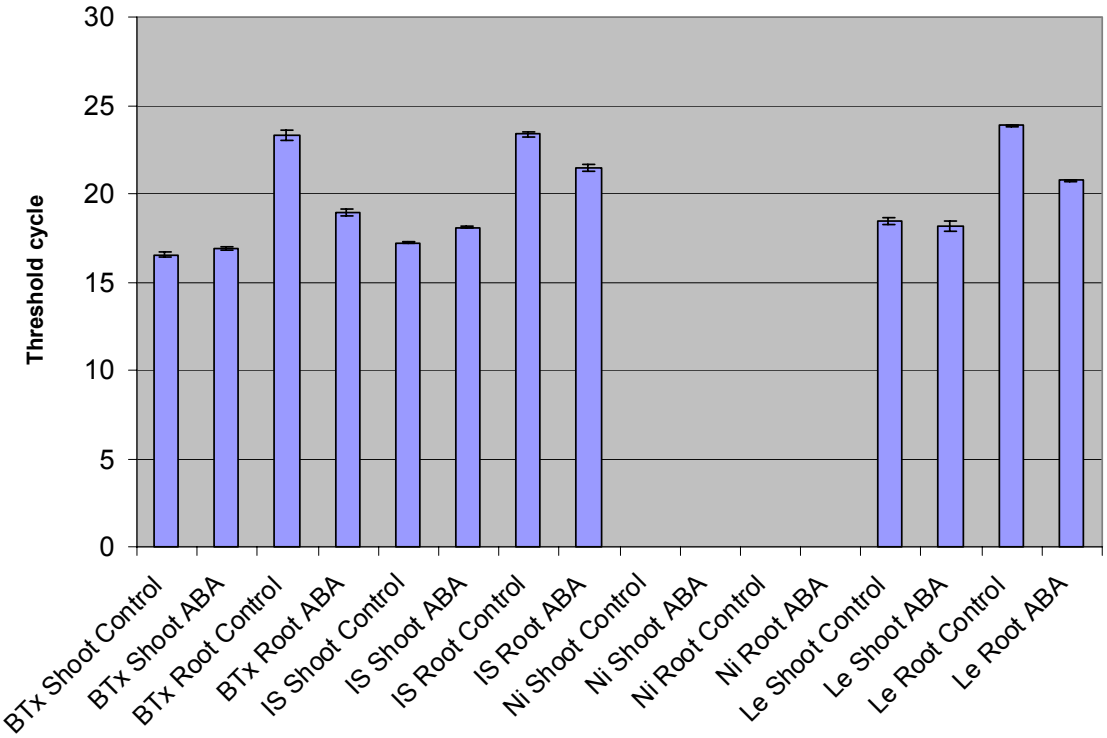

Figure 14. Expression level of predicted gene 181g10_10 under control and ABA treatment. The primers designed for Nipponbare failed to meet the primer efficiency criteria, no RT-PCR data is available and the plot is left as empty. SD range is calculated by upper error bar and lower error bar on top of each column. BTx= BTx623, IS= IS3620C, Ni= Nipponbare, Le= Lemont

Rice and sorghum are very different grasses and thus a direct comparison of the level of gene expression in control tissues between these two species isn't valid. Therefore, the approach taken in the present research was to examine whether the loss of gene colinearity resulted in differential gene expression following some type of treatment known to alter gene expression. It is well known that ABA is involved in the regulation of hundreds of genes in sorghum (Buchanan et al., 2005). It was not known whether any of the genes chosen for this study were responsive to ABA treatment, however, since ABA is induced by many different abiotic stress conditions and it is involved in numerous developmental processes it was deemed a good choice as an initial treatment. Two sorghum genotypes and two rice genotypes were also included in the analysis, however, it appears that there were no genotype-specific affects on gene expression. Twelve colinear and 5 non-colinear genes were used for final RT-PCR analysis although the initial intent was to examine a similar number of colinear and non-colinear genes. The colinear genes selected were from the fully sequenced sorghum BAC clones, 181g10 and 82G24 (Table VI and VII) since their colinearity to rice could easily be established. Non-colinear genes were selected from the sequences obtained from the 13 BAC pools used for the 1.5× sequence skim analysis (supplemental Table XII) and predicted genes like transcription factors were preferred because they most likely would show differential expression following ABA treatment. Unfortunately it was difficult to find a complete set of primers for 12 non-colinear genes that had similar primer efficiency for all 4 sorghum and rice varieties. While it was possible to identify 12 colinear genes and design efficient primers for these genes, fewer non-colinear genes

were available since in most cases it was difficult to positively identify the sorghum-rice ortholog pairs.

In this research, no direct link between colinearity and gene expression pattern was identified for the genes examined. It is possible that there is no relationship between the loss of gene colinearity and changes in expression pattern between species. Alternatively, it may be that the proper conditions required to detect a difference in gene expression between sorghum and rice were not selected in this study. In future research, sequence information adjacent to the genes, especially the upstream sequence, may help identify better candidate genes for study. Additionally, qRT-PCR analysis for a larger number of genes and more treatments will be helpful to identify the possible relationship between loss of colinearity and expression pattern change.

CHAPTER VI

CONCLUSION

Integrated genetic and physical maps are extremely important for map-based gene cloning, comparative genome analysis, and genome sequencing. Using the rice chromosome 1 physical map as a reference, a minimal tiling path for the euchromatic arms of sorghum chromosome 3 was constructed using a combination of six-dimensional (6D) BAC DNA pooling, EST-STS PCR screening, AFLP analysis, and HICF fingerprinting. Sorghum ESTs were used to design primers for screening BAC DNA pools derived from the parents of the TAMU-ARS RI mapping population (BTx623 and IS3620C).

Six hundred and twelve single- or low-copy sorghum ESTs were selected for EST-STS primer design. Four hundred and sixty-six primer sets produced strong PCR signals in the BAC pools that could be easily analyzed, of which 447 primer sets gave positive signals in the BTx623 BAC pools and 452 in the IS3620C BAC pools. In total, 2080 sorghum BAC clones were identified by these primer sets, of which 1093 BAC clones are derived from BTx623 and 987 are derived from IS3620C. On average, each primer set identified ~2.45 BAC clones from the BTx623 BAC pools and ~2.18 from the IS3620C BAC pools.

In a previous study, AFLP technology was combined with the BAC DNA pooling strategy to allow overlapping BAC clones to be identified while simultaneously linking

the genetic and physical maps (Klein et al., 2000). Using this technique, 114 AFLP genetic markers were linked to BTx623 BAC clones within the sorghum physical map (Klein et al., 2000). In the present study, 94 AFLP markers (59 located within the euchromatic arms and 35 located within the heterochromatin) were scored for their presence in the IS3620C BAC DNA pools providing additional linkage between the genetic and physical maps. In total, the minimal tiling path of sorghum chromosome 3 contains 290 genetic markers (AFLPs, SSRs, Indels, and EST-SSRs).

Sorghum BAC clones identified by EST-STS PCR screening and AFLP technology were fingerprinted by the HICF method and contigs were constructed using the program FPC V8.2 (Soderlund et al., 1997). Twenty-three contigs were created to represent the minimal tiling path of the euchromatic arms. The average contig size is 2.50 Mb, while the largest is 13.78 Mb and the smallest is 210 kb. These 23 contigs represent 57.56 Mb of the euchromatic regions of sorghum chromosome 3, of which 25.79 Mb represents the short arm and 31.77 Mb represents the long arm. Gap sizes between these contigs were estimated based on the size of the colinear region in rice and all gaps are estimated to be no more than 380 kb except the gap which resides at the junction of the major inversion within the short arm, the size of which could not be determined.

Contigs within the heterochromatic region of sorghum chromosome 3 were also constructed with limited success. Thirty-six primer sets produced strong and analyzable PCR signals. Of these 9 EST-STS markers collocated in contigs containing chromosome 3 genetic markers located within the heterochromatin. In addition, another 8 EST-STS markers collocated in 2 contigs (one contig contained 3 EST-STS markers

and the other contained 5 EST-STS markers) although these contigs did not contain any genetic markers. These results indicate that these blocks of genes have remained colinear with respect to their order in rice although it can not be determined at present whether these gene blocks are located on the homeologous rice and sorghum chromosomes. Analysis of BAC clones identified with 15 of the EST-STSs thought to be located within the sorghum chromosome 3 heterochromatin indicated a loss of colinearity of these genes between sorghum and rice. The data suggests that 5 genes have relocated to other regions within sorghum chromosome 3 (i.e., movement from heterochromatin to euchromatin) whereas the remaining 10 genes were relocated to other sorghum chromosomes. Data on the remaining 4 EST-STS markers was inconclusive as these markers were individually localized to unmapped BAC contigs. As mentioned above, AFLP analysis in the BAC DNA pools was used to link 35 genetic markers to contigs within the heterochromatin. In total, 21 contigs that span the large pericentromeric heterochromatin were constructed representing 16.57 Mb of DNA.

Kim et al. (2005a) used FISH analysis to estimate that sorghum chromosome 3 contains a total of 89.9 Mb, of which 51.7 Mb is from the euchromatic regions and the remaining 38.2 Mb resides in the heterochromatin. The present estimates of the DNA represented in the sorghum chromosome 3 minimal tiling path (both euchromatin and heterochromatin) are based on HICF analysis of 15 fully sequenced sorghum BAC clones where the average number of ~75- to 500-bp DNA bands detected by HICF per megabase pair of DNA was calculated. This analysis is likely an overestimate of the actual DNA content represented in the minimal tiling path because the current HICF

map contains BAC clones from two highly divergent sorghum genotypes (BTx623 and IS3620C) and due to the inherent limitations associated with BAC DNA fingerprinting (Nelson et al., 2005). In any case, even if the present results are overestimating the DNA content by 10-15% (57.46 Mb in euchromatin in present study vs. 51.7 Mb based on FISH analysis), the present results suggest that the minimal tiling path represents nearly 100% of the DNA from the euchromatic arms and ~35-40% of the DNA within the heterochromatin.

Macrocolinearity between sorghum chromosome 3 and rice chromosome 1 was examined based on the mapped EST-STS markers. Three hundred and eighty-eight out of 456 mapped EST-STS markers were colinear between sorghum chromosome 3 and rice chromosome 1 indicating a level of colinearity of ~85%. Gene translocations, duplications, and inversions relative to rice chromosome 1 were detected within sorghum chromosome 3.

Construction of the integrated genetic and physical map of the euchromatic arms of sorghum chromosome 3 allowed the average rates of recombination to be determined. The rate of recombination of euchromatic arm varies from 0.068 Mb/cM to 0.447 Mb/cM with the average of 0.204 Mb/cM indicating the existence of recombination cold and hot spots. The rate of recombination estimated here differs by ~1.25-fold to the estimate by Kim et al (2005a) which may be caused by the overestimate of the actual DNA content represented in the minimal tiling path of sorghum chromosome 3 that is associated with the inherent limitations of BAC DNA fingerprinting as described previously (Nelson et al., 2005). However, the variation in the rate of recombination

across sorghum chromosome 3 observed in the present study is similar to that observed by Kim (J.S. Kim, personal communication).

Microcolinearity between sorghum chromosome 3 and rice chromosome 1 was examined at two different levels. First, a region of sorghum chromosome 3 represented by 13 overlapping BAC pools was compared to the 5.1 Mb orthologous region of rice chromosome 1 by sequence skimming the sheared BAC pool libraries to a depth of ~1.5× and comparing the resulting sorghum sequences to the TIGR gene models from the 12 rice pseudomolecules. Second, 2 sorghum BAC clones from this region were completely sequenced and annotated to further detail microcolinearity in this region.

Seven hundred and twenty-nine skim sequences from 13 BAC pools were identified as genic, and 472 of them were colinear relative to rice based on the current TIGR gene models. From this analysis, ~62% of the gene sequences identified were colinear between sorghum and rice, although the colinearity of genes within individual pools varied from ~54% to ~78%. The 257 non-colinear sorghum gene sequences had hits to all 12 rice pseudomolecules (BLASTX analysis, e-value > e^{-10}). Most of these non-colinear gene sequences were distributed randomly across the rice genome except for a region on rice chromosome 5. A number of sorghum gene sequences from pool 1 to pool 8 aligned to the rice chromosome 5 pseudomolecule from ~19.5 to 23.5 Mb which likely reflects the segmental homology between rice chromosomes 1 and 5 resulting from the whole genome duplication of rice predating the origination of the grasses (Yu et al., 2005).

Two sorghum BAC clones, 82G24 (IS3620C genotype) and 181g10 (BTx623 genotype), from the region covered by the sorghum BAC pools were fully sequenced and annotated. BAC 82G24 is ~200 kb and BAC 181g10 is ~110 kb. Twenty genes were predicted to be encoded in BAC 82G24, averaging one gene per 10.0 kb, while 15 genes were predicted to be encoded in BAC 181g10, averaging one gene per 7.3 kb. The corresponding homeologous rice genomic sequences were pulled out from the rice chromosome 1 pseudomolecule, AP008207, and the rice genes were also predicted. Approximately 65% colinearity was detected between sorghum BAC 82G24 and ~59% colinearity was detected between sorghum BAC 181g10 and the corresponding regions of the rice chromosome 1 pseudomolecule. Regions of microcolinearity reflected gene-coding sequences and the regions immediately flanking these genes, whereas loss of colinearity reflected missing genic sequences and intergenic spaces. Small-scale gene rearrangements including gene duplication, gene loss, gene translocation, and gene inversion were detected.

While macrocolinearity analysis indicated ~85% colinearity between sorghum chromosome 3 and rice chromosome 1, microcolinearity analysis indicated a level of colinearity between ~59% to ~65%. The ability to detect small-scale rearrangements including segmental duplication, gene amplification, gene loss, and gene translocation by analysis of microcolinearity at the DNA sequence level may be the reason for the difference in colinearity observed at these two levels. In addition, it is likely that the EST-STS marker strategy is overestimating the true level of colinearity between sorghum and rice due to the preselection of single- or low-copy ESTs for the analysis.

Although the analysis of macrocolinearity and microcolinearity between sorghum and rice indicates a high degree of colinearity between these two cereals that diverged from a common ancestor more than 50 MYA (Doebley et al., 1990), many examples of loss of gene colinearity were detected. This loss of colinearity leads to the question of whether translocation of genes within the genome has led to an associated change in the transcriptional regulation of these genes. To address this question, qRT-PCR analysis of a number of colinear and non-colinear genes within the analyzed regions of sorghum and rice were carried out in control and ABA-treated root and shoots. Although expression pattern differences were detected, some were species specific (rice vs. sorghum) and some were tissue specific (shoot vs. root), none of the expression pattern changes could be directly linked to the loss of colinearity. These results suggest that there may be no change of expression pattern related to the movement of these genes, or if there is, the conditions used in the present study were not sufficient to detect these changes. qRT-PCR analysis for more genes and more treatments may help to find a possible relationship between loss of colinearity and expression pattern changes in future research.

LITERATURE CITED

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proceedings of the National Academy of Sciences USA **90:** 7980-7984

Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, Bosch NVD, Pot J, Kuiper MTR (1998) Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant Journal **14:** 259-271

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter **9:** 208-219

Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N (1997) Human BAC library: construction and rapid screening. Gene **191:** 69-79

Avramova Z, Tikhonov A, SanMiguel P, Jin Y-K, Liu C-N, Woo S-S, Wing RA, Bennetzen JL (1996) Gene identification in a complex chromosomal continuum by local genomic cross-referencing. Plant Journal **10:** 1163-1168

Barkan A, Martienssen RA (1991) Inactivation of maize transposon Mu suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. Proceedings of the National Academy of Sciences USA **88:** 3502-3506

Barrero JM, Piqueras P, Gonzalez-Guzman M, Serrano R, Rodriguez PL, Ponce MR, Micol JL (2005) A mutational analysis of the ABA1 gene of *Arabidopsis*

thaliana highlights the involvement of ABA in vegetative development. J Exp Bot **56:** 2071-2083

Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, et al (2005) Sorghum genome sequencing by methylation filtration. Plos Biology **3:** 103-115

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell **12:** 1021-1029

Bennetzen JL, Coleman C, Liu RY, Ma JX, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Current Opinion in Plant Biology **7:** 732-736

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends in Genetics **9:** 259-261

Bennetzen JL, Freeling M (1997) The unified grass genome - synergy in synteny. PCR Methods & Applications **7:** 301-306

Bennetzen JL, Ramakrishna W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Molecular Biology **48:** 821-827

Bhattramakki D, Dong JM, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome **43:** 988-1002

Boivin K, Deu M, Rami JF, Trouche G, Hamon P (1999) Towards a saturated sorghum map using RFLP and AFLP markers. Theoretical & Applied Genetics **98:** 320-328

Borrell AK, Bidinger FR, Sunitha K (1999) Stay-green trait associated with yield in recombinant inbred sorghum lines varying in rate of leaf senescence. International Sorghum & Millets Newsletter **40:** 31-34

Borrell AK, Hammer GL (2000a) Nitrogen dynamics and the physiological basis of stay-green in sorghum. Crop Science **40:** 1295-1307

Borrell AK, Hammer GL, Henzell RG (2000b) Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. Crop Science **40:** 1037-1048

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, et al (2003) A high-density genetic recombination map of sequence-tagged sites for *Sorghum*, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics **165:** 367-386

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, et al (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proceedings of the National Academy of Sciences, USA **102:** 13206-13211

Brenner S, Livak KJ (1989) DNA fingerprinting by sampled sequencing. Proceedings of the National Academy of Sciences, USA **86:** 8902-8906

Buchanan CD, Lim S, Salzman RA, Kagiampakis I, Morishige DT, Weers B, Klein RR, Pratt LH, Cordonnier-Pratt M-M, Klein PE, Mullet JE (2005) *Sorghum bicolor's* transcriptome response to dehydration, high salinity and ABA. Plant Molecular Biology **58:** 699-720

Cao Y, Kang HL, Xu X, Wang M, Dho SH, et al (1999) A 12-Mb complete coverage BAC contig map in human chromosome 16p13.1-p11.2. Genome Research **9:** 763-774

Chen M, SanMiguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proceedings of the National Academy of Sciences, USA **94:** 3431-3435

Cheng WH, Endo A, Zhou L, Penney J, Chen HC, Arroyo A, Leon P, Nambara E, Asami T, Seo M, Koshiba T, Sheen J (2002) A unique short-chain dehydrogenase/reductase in Arabidopsis glucose signaling and abscisic acid biosynthesis and functions. Plant Cell **14:** 2723-2743

Childs KL, Klein RR, Klein PE, Morishige DT, Mullet JE (2001) Mapping genes on an integrated sorghum genetic and physical map using cDNA selection technology. The Plant Journal **27:** 243-255

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* x *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of Sorghum chromosomes or chromosomal segments. Theoretical & Applied Genetics **87:** 925-933

Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. Proceedings of the National Academy of Sciences, USA **96:** 15330-15335

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Molecular Biology **35:** 3-15

Ding Y, Johnson MD, Chen WQ, Wong D, Chen Y-J, Benson SC, Lam JY, Kim Y-M, Shizuya H (2001) Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using Type IIS restriction endonucleases. Genomics **74:** 142-154

Doebley J, Durbin M, Golenberg EM, Clegg MT, Ma DP (1990) Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (Gramineae). Evolution **44:** 1097-1108

Doggett H (1988) Sorghum, 2nd ed. John Wiley, New York, NY

Draye X, Lin YR, Qian XY, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiology **125:** 1325-1341

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan LL, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiology **125:** 1342-1353

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glazmann JC, Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theoretical & Applied Genetics **94:** 409-418

Duncan RR, Bockholt AJ, Miller FR (1981) Descriptive comparison of senescent and non-senescent sorghum genotypes. Agronomy Journal **73:** 849-853

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proceedings of the National Academy of Sciences, USA **96:** 8265-8270

Feuillet C, Penger A, Gellner K, Mast A, Keller B (2001) Molecular evolution of receptor-like kinase genes in hexaploid wheat. Independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. Plant Physiology **125:** 1304-1313

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell **14:** S15-S45

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science **282:** 656-659

Gnansounou E, Dauriat A, Wyman CE (2005) Refining sweet sorghum to ethanol and sugar: economic trade-offs in the context of North China. Bioresource Technology **96:** 985-1002

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proceedings of the National Academy of Sciences, USA **87:** 1213-1217

Gregory SG, Howell GR, Bentley DR (1997) Genome mapping by fluorescent fingerprinting. Genome Research **7:** 1162-1168

Hardison RC (2000) Conserved noncoding sequences are reliable guides to regulatory elements. Trends in Genetics **16:** 369-372

Howe A, Sato S, Dweikat I, Fromm M, Clemente T (2006) Rapid and reproducible Agrobacterium-mediated transformation of sorghum. Plant Cell Rep **25:** 784-791

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proceedings of the National Academy of Sciences, USA **87:** 4251-4255

Islam-Faridi MN, Childs KL, Klein PE, Hodnett G, Menz MA, Klein RR, Rooney WL, Mullet JE, Stelly DM, Price HJ (2002) A molecular cytogenetic map of sorghum chromosome *1*: Fluorescence *in situ* hybridization analysis with mapped bacterial artificial chromosomes. Genetics **161:** 345-353

Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M (2002) Utility and distribution of conserved noncoding sequences in the grasses. Proceedings of the National Academy of Sciences, USA **99:** 6147-6151

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiology **125:** 1198-1205

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from Japonica rice. Science **301:** 376-379

Kim J, Gordon L, Dehal P, Badri H, Christensen M, Groza M, Ha C, Hammond S, Vargas M, Wehri E, Wagner M, Olsen A, Stubbs L (2001) Homology-driven assembly of a sequence-ready mouse BAC contig map spanning regions related

to the 46-Mb gene-rich euchromatic segments of human chromosome 19. Genomics **74:** 129-141

Kim JS, Childs KL, Faridi NI, Menz MA, Klein RR, Klein PE, Price HJ, Mullet JE, Stelly DM (2002) Integrated karyotyping of sorghum by *in situ* hybridization of landed BACs. Genome **45:** 402-412

Kim JS, Islam-Faridi MN, Klein PE, Stelly DM, Price HJ, Klein RR, Mullet JE (2005a) Comprehensive molecular cytogenetic analysis of sorghum genome architecture; distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics **171:** 1963-1976

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005b) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics **169:** 1169-1173

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005c) Molecular cytogenetic maps of sorghum linkage groups 2 and 8. Genetics **169:** 955-965

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Research **10:** 789-807

Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. The Plant Journal **34:** 605-621

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theoretical & Applied Genetics **111:** 994-1012

Klein RR, Morishige DT, Klein PE, Dong J, Mullet JE (1998) High throughput BAC DNA isolation for physical map construction of sorghum (*Sorghum bicolor*). Plant Molecular Biology Reporter **16:** 351-364

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theoretical & Applied Genetics **101:** 438-448

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences USA **95:** 370-375

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics **82:** 378-389

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155

Magalhães JV, Garvin DF, Wang Y, Sorrells ME, Klein PE, Schaffert RE, Li L, Kochian LV (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics **167:** 1905-1914

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, et al (1997) High throughput fingerprint analysis of large-insert clones. Genome Research **7:** 1072-1084

McBee GG (1984) Relation of senescence, nonsenescence, and kernel maturity to carbohydrate metabolism in sorghum. *In* LK Mughogho, ed, Sorghum Root and Stalk Diseases, a Criticl Review. Proceedings of the Consultative Group Discussion of Research Needs and Strategies for Control of Sorghum Root and Stalk Diseases. ICRISAT, Bellagio, Italy, pp 119-129

Melake Berhan A, Hulbert SH, Butler LG, Bennetzen JL (1993) Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theoretical & Applied Genetics **86:** 598-604

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP$^{(R)}$, RFLP and SSR markers. Plant Molecular Biology **48:** 483-499

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell **10:** 1817-1832

Moore G, Devos KM, Wang Z, Gale MD (1995) Grasses, line up and form a circle. Current Biology **5:** 737-739

Nelson WM, Bharti AK, Butler E, Wei FS, Fuks G, Kim H, Wing RA, Messing J, Soderlund C (2005) Whole-genome validation of high-information-content fingerprinting. Plant Physiology **139:** 27-38

Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA (2003) Structure and evolution of cereal genomes. Current Opinion in Genetics & Development **13:** 644-650

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, et al (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature **400:** 256-261

Peng Y, Schertz KF, Cartinhour S, Hart GE (1999) Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breeding **118:** 225-235

Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. Genome **37:** 236-243

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus *Sorghum* (Poaceae). Annals of Botany **95:** 219-227

Qi X, Stam P, Lindhout P (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. Theoretical & Applied Genetics **96:** 376-384

Rosenow DT, Clark LE (1981) Drought tolerance in sorghum. *In* HD Loden, D Wilkinson, eds, Proceedings of the 36th Annual Corn and Sorghum Industry Research Conference. American Seed Trade Association, Chicago, IL., pp 18-30

Rosenow DT, Quisenberry JE, Wendt CW, Clark LE (1983) Drought tolerant sorghum and cotton germplasm. Agricultural Water Management **7:** 207-222

Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Idonuma A, Watanabe W, Shimizu Y, Horiuchi I, Matsumoto T, Sasaki T, Higo K (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Research **30:** 98-102

Salzman RA, Brady JA, Finlayson SA, Buchanan CD, Summer EJ, et al (2005) Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses. Plant Physiology **138:** 352-368

Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). Plant Molecular Biology **48:** 713-726

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, et al (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765-768

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Functional & Integrative Genomics **2:** 70-80

Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr Opin Plant Biol **3:** 138-141

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, et al (2002) The genome sequence and structure of rice chromosome 1. Nature **420:** 312-316

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker--a web server for aligning two genomic DNA sequences. Genome Research **10:** 577-586

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. Computer Applications in the Biosciences **13:** 523-535

Song R, Llaca V, Linton E, Messing J (2001) Sequence, regulation, and evolution of the maize 22-kD á zein gene family. Genome Research **11:** 1817-1825

Song R, Llaca V, Messing J (2002) Mosaic organization of orthologous sequences in grass genomes. Genome Research **12:** 1549-1555

Sorghum Genomics Planning Workshop Participants (2005) Toward sequencing the sorghum genome: a US National Science Foundation-sponsored workshop report. Plant Physiology **138:** 1898-1902

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell **12:** 381-391

Thomas JW, Summers TJ, Lee-Lin SQ, Maduro Braden VV, Idol JR, et al (2000) Comparative genome mapping in the sequence-based era: early experience with human chromosome 7. Genome Research **10:** 624-633

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999) Colinearity and its exceptions in orthologous *adh* regions of

maize and sorghum. Proceedings of the National Academy of Sciences, USA **96:** 7409-7414

Ventelon M, Deu M, Garsmeur O, Doligez A, Ghesquiere A, Lorieux M, Rami JF, Glaszmann JC, Grivet L (2001) A direct comparison between the genetic maps of sorghum and rice. Theoretical & Applied Genetics **102:** 379-386

Vollrath D, Jaramillo-Babb VL (1999) A sequence-ready BAC clone contig of a 2.2-Mb segment of human chromosome 1q24. Genome Research **9:** 150-157

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, et al (1999) Two high-density AFLP$^{(R)}$ linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theoretical & Applied Genetics **99:** 921-935

Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, et al (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant Journal **27:** 581-590

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics **132:** 1119-1130

Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics **153:** 453-473

Woo SS, Jiang JM, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Research **22:** 4922-4931

Woodfin CA, Rosenow DT, Clark LE (1998) Association between the stay-green trait and lodging resistance in sorghum. *In* Agronomy Abstracts. Agronomy Society of America, ASA, Madison, WI, p 102

Wu JZ, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, et al (2004) Composition and structure of the centromeric region of rice chromosome 8. Plant Cell **16:** 967-976

Xu W, Rosenow DT, Nguyen HT (2000a) Stay green trait in grain sorghum: relationship between visual rating and leaf chlorophyll concentration. Plant Breeding **119:** 365-367

Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000b) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome **43:** 461-469

Xu WW, Crasta O, Rosenow DT, Mullet JE, Nguyen HT (1994) Progress toward molecular mapping of drought resistance traits in sorghum. International Sorghum & Millets Newsletter **35:** 91-92

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Sciences, USA **100:** 6263-6268

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. Theoretical & Applied Genetics **99:** 785-790

Yu J, Hu S, Wang J, Wong GK-S, Li S, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79-92

Yu J, Wang J, Lin W, Li SG, Li H, et al (2005) The genomes of *Oryza sativa*: a history of duplications. Plos Biology **3:** 266-281

Zhao ZY, Cai TS, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D (2000) Agrobacterium-mediated sorghum transformation. Plant Molecular Biology **44:** 789-798

Zhu H, Blackmon BP, Sasinowski M, Dean RA (1999) Physical map and organization of chromosome 7 in the rice blast fungus, *Magnaporthe grisea*. Genome Research **9:** 739-750

Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol **53:** 247-273

APPENDIX

Table X. Genetic markers linked to the sorghum chromosome 3 minimal tiling path

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa2725	22.23	off-end			
txa2976	24.67	off-end			
txp456	36.3	off-end		457	CW276008
txa6000	21.67	off-end		457	
txa3850	32.67	off-end		457	
txa3395	**25.21**	**framework**	**0.0**	**457**	
txa2580	**30.85**	**framework**	**1.9**	**457**	
txs333	29.03	region			
txp496	32.24	region			CD203723
txa3461	**28.8**	**framework**	**4.3**		
txa6154	**36.53**	**framework**	**4.7**		
txp494	36.93	unique		457	BG933631
txa2105	**7.32**	**framework**	**6.6**	**457**	
SDB052	35.05	region		457	
txp457	34.5	region		8642	CD234238
umc121	**25.45**	**framework**	**9.0**		
txa3159	28.52	region		8642	
txa2224	38.43	unique		8642	
TS452	34.45	region		8642	
txa6263	32.54	region		8642	
txa2079	31.72	unique		8642	
txa261	32.97	unique		8642	
txs31	32.24	unique		8642	
txs1178	**28.17**	**framework**	**11.9**		
txp518	34.45	unique		8642	
txa2849	29.69	region		8642	
txp266	32.83	unique			
TS483	37.83	unique			
txp228	25.04	region			
txa4134	**22.77**	**framework**	**15.2**	**8642**	
txp454	31.5	unique		8642	CL149712
txa375	33.26	region		8642	
txa376	**26.8**	**framework**	**16.7**	**8642**	
txp492	27.96	unique		8642	CF484286
txa552	24.34	unique			
umc124.1	**23.67**	**framework**	**20.3**		
txa3498	**32.67**	**framework**	**20.8**		
txa6160	**29.89**	**framework**	**22.4**	**8642**	
txa6272	**27.8**	**framework**	**25.2**		
txp491	23.53	unique		8642	CW095430
txa2734	**26.01**	**framework**	**28.6**	**8642**	
ra2	25.89	unique			
txa2598	19.8	unique		8642	
txa610	30.18	region		498	
txa2614	**29.32**	**framework**	**31.6**	**498**	
txa4216	30.78	region		498	
txa4115	34.45	region			
txa2328	33.12	unique		498	

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa6004	35.73	unique		498	
txs1092	**-**	**framework**	**33.3**	**498**	
txa6359	32.54	unique		498	
txa6292	29.89	unique			
txa3590	19.72	unique		30	
txa3569	23.27	region			
txa3838	24.49	region		8735	
txa2097	34.75	region			
txa6291	30.18	unique		23	
txa3003	**21.66**	**framework**	**37.4**	**23**	
SDB050	33.42	unique		23	
txp489	35.94	unique		23	BG049321
txa6025	33.12	region		23	
txa2096	**29.03**	**framework**	**39.4**	**23**	
txa2714	**29.59**	**framework**	**41.4**		
txp451	19.4	unique		23	CW247904
txp452	37.5	region		23	CW265582
txa4126	29.03	unique			
txa3848	**29.88**	**framework**	**43.4**	**23**	
txp215	16.67	unique		23	
txa6085	24.13	unique		23	
txa6218	28.82	unique			
txp488	19.56	unique		23	CW083960
txa3552	**19.8**	**framework**	**49.4**	**516**	
txa2671	31.9	region		516	
txa2546	**28.52**	**framework**	**51.8**	**516**	
txa267	**28.53**	**framework**	**53.3**	**516**	
txa2822	34.83	unique		516	
txp423	33.56	unique		516	C2_2544
SDB048	39.93	unique		516	
txa2967	32.08	unique		516	
txa6042	34.59	unique		516	
txa4102	29.32	unique			
txa2165	31.07	unique			
txa2107	28.8	unique			
txa2520	22.71	region			
txa29	**27.03**	**framework**	**55.3**	**516**	
txa30	**36.83**	**framework**	**55.7**	**516**	
txa2336	28.53	unique		301	
txs1053	31.66	unique			
txa6033	34.15	unique			
BH245389	36.23	region		9539	pSB0485
txa3021	**28.24**	**framework**	**58.1**		
txa2260	**39.33**	**framework**	**58.1**	**9539**	
txp485	26.2	unique		9539	CW485748
txp554	26.77	unique		9539	
txa3837	**25.21**	**framework**	**61.6**	**8664**	
txa3839	17.46	region		8664	
txa2996	24.4	unique			
txa4173	22.45	region		1232	
txa2899	**24.13**	**framework**	**65.1**		

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa434	26.57	unique			
txa6027	33.46	region		1232	
txa3485	30.27	region		1232	
txa556	22.23	unique		283	
txa2746	28.46	region			
txa2634	27.4	region		9058	
txs578	**-**	**framework**	**68.3**		
txa234	26.77	unique		1739	
txa3781	22.5	unique		1739	
txa2246	24.76	unique		1739	
txp500	33.71	unique		1739	CW383070
txa2989	24.76	unique			
isu148	**24.13**	**framework**	**71.8**		
txa3450	28.52	unique		1739	
txa3039	31.95	region		1739	
txs503	30.19	unique		1739	
txa366	29.69	region			
txa3718	20.61	region			
txa2171	**25.17**	**framework**	**74.4**	**1739**	
txa3583	25.45	region		1739	
txa3267	27.05	region		1739	
txa3116	27.33	region		1739	
txa3603	28.17	region		1739	
txa3717	29.89	unique			
txa402	29.11	region			
txa2273	32.24	region			
BH246082	35.64	unique			pSB1693.2
umc152.1	**26.75**	**framework**	**76.1**		
txa2116	25.49	region		608	
txa476	33.42	unique		608	
txa6050	30.27	region		508	
txa2646	27.61	region		508	
txa3313	30.16	unique		508	
txa3392	31.61	region		508	
txa4027	32.24	unique		508	
txa2904	36.53	unique		508	
txp461	35.7	unique		508	CF485333
txa2670	36.23	unique		508	
txa2109	32.73	unique		508	
txa3947	35.73	unique		508	
txa2058	26.77	unique			
txa2305	36.23	unique		772	
txa205	35.94	unique		772	
txa2019	31.95	unique		772	
txa2030	33.86	unique		772	
txa3375	36.83	unique		772	
txa2015	37.83	unique			
txa484	34.83	unique			
txa535	24.67	unique			
txa576	37.83	unique		5995	
txa197	35.05	unique		3826	

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa186	28.18	unique			
txa450	26.69	unique		9107	
txa418	28.82	unique		5	
txa3740	21.41	region		5	
txa6364	35.64	unique		1125	
txa3637	39.63	unique		1125	
txa6352	32.3	unique		219	
txa2745	35.05	unique		252	
txa6001	37.53	unique		252	
txa4174	33.71	unique			
txa3852	31.48	unique			
txa3780	23.86	unique			
txa3753	25.17	unique			
txa3668	36.23	unique			
txa3656	35.05	unique		8769	
txa3646	37.83	unique			
txa6161	39.93	unique		1370	
txa3618	30.75	unique		1370	
txa3591	27.8	unique		689	
txa3571	32.88	unique			
txa3489	36.03	unique			
umc10	33.71	unique			
txa3462	39.63	unique			
txa3742	22.45	region		1238	
txa3460	36.23	unique		1238	
txa3417	31.14	unique		singleton	
txa3410	31.61	unique		10468	
txa3250	30.56	unique		8685	
txa3246	30.46	unique		8685	
txa3220	31.61	unique			
txa3139	27.12	unique			
txa3122	31.95	unique		502	
txa3823	33.71	region		10013	
txa3030	29.89	unique			
txa2969	33.56	unique			
txa2917	35.64	unique			
txa2817	30.46	unique			
txa2779	36.93	unique			
txa2774	28.46	unique		8875	
txa2759	20.42	unique			
txs1438	37.83	unique			
txa2642	36.33	unique			
txa2586	32.88	unique		896	
txa2529	33.71	unique			
txs1162	37.83	unique		1181	
txa2205	31.14	unique			
txa2137	26.42	unique			
txa6253	35.64	region			
txa4192	24.13	region			
txa3703	31.43	region			
txa3606	31.04	region			

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa3835	33.42	region		3708	
txa3212	32.3	region		10545	
txa2825	30.16	region		8909	
txa2805	33.26	region			
TS487	32.54	region			
txa6157	38.43	unique		5929	
txa2794	36.53	unique		5929	
txa3112	31.43	region		5929	
rz244	**23.67**	**framework**	**79.5**		
txp33	20.83	region			
txa6276	37.83	unique		5291	
txa3701	28.18	region		5291	
cdo920	**25.36**	**framework**	**81.9**		
txa3688	25.86	region		77	
Corn023	34.15	region		77	
txa6175	31.19	unique		77	
txa6031	32.3	unique		77	
txa4082	33.86	unique			
txa2768	31.66	region		77	
gap236	25.36	region		77	
isu114	**25.08**	**framework**	**84.3**		
txa3350	37.53	unique		77	
txa4147	36.33	unique		77	
txa330	33.71	region		77	
txp205	20.67	unique		77	
txa3270	18.96	region		77	
txa6335	36.63	unique			
txa3493	29.88	unique		10664	
txa3836	35.05	region			
txa3102	31.32	region			
txa2880	25.45	region			
txa3261	**24.21**	**framework**	**88.2**		
txa2923	32.01	region			
txp31	22	unique		385	
txp507	30.27	region		385	CL195826
txa4054	**28.18**	**framework**	**90.6**	**385**	
txp336	22.77	unique			
txa3378	21.19	region			
txp183	**21.03**	**framework**	**94.0**		
txa3294	26.42	unique		385	
txp444	28.5	region		385	BG355482
txa2735	**18.96**	**framework**	**98.8**	**385**	
txp543	36.63	region		385	6878.Contig2
txa2320	29.89	region		385	
txa242	25.31	unique		385	
txa2826	23.24	unique			
txa2769	27.4	unique			
txa2354	30.18	unique			
txa2607	**25.31**	**framework**	**102.6**		
txp544	36.93	unique		385	43216.Contig1
txa4106	**28.14**	**framework**	**104.8**	**385**	

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa4194	31.95	unique		385	
txs584	-	region		385	
txa6099	30.85	unique			
txp503	33.86	unique		385	AF010283
txa6309	**32.54**	**framework**	**105.9**	**385**	
umc93	**27.61**	**framework**	**108.2**		
umc63	**36.33**	**framework**	**108.2**		
txp120	**27.68**	**framework**	**110.9**	**385**	
TS120N	35.94	region		385	
txp435	35.05	region		385	BM317672
txa6219	32.24	unique		385	
txa2595	**37.53**	**framework**	**111.6**	**385**	
txp546	23.92	region			14807.Contig3
txp511	16.34	unique		385	CL159968
txa2074	**34.59**	**framework**	**111.6**	**385**	
txp545	29.31	unique		385	24629.Contig1
txp2	27.68	unique		385	
txp529	27.66	unique		385	
cdo1160	**15.62**	**framework**	**118.4**		
txp436	20.61	region		4076	BG556153
txp231	14.76	unique		4076	
txp59	18.75	unique		4076	
txa2110	**20.35**	**framework**	**125.2**	**4076**	
txa389	**35.43**	**framework**	**125.2**	**4076**	
txa6188	29.69	region		4076	
txa3940	32.73	region		4076	
txa3731	23.58	region		4076	
txa255	26.2	region			
txa2233	18.63	region		4611	
txa6212	32.24	region		4611	
txa3579	34.3	unique		4611	
txa6155	34.3	unique		4611	
txa2506	31.48	region			
CS032	31.66	region		4611	
txa4131	31.04	unique		4611	
txs1175	**28.17**	**framework**	**127.5**		
isu74.2	**30.19**	**framework**	**128.6**		
TS431B	29.02	region			
cdo470	**26.51**	**framework**	**130.1**		
txp218	19.39	unique			
bcd828	29.9	unique			
txa337	32.43	region		4611	
txa3295	21.96	unique		4611	
txa54	34.23	region		4611	
txa4019	**28.25**	**framework**	**133.3**	**4611**	
txa38	37.83	region		4611	
txa69	29.4	region		4611	
txa2197	39.63	unique		4611	
isu121	**34.53**	**framework**	**133.3**		
txa593	20.34	region		94	
txa594	25.51	unique		94	

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
umc17	**36.33**	**framework**	**134.7**		
txa392	26.01	region		94	
txa3915	33.86	region		94	
txp114	27.03	unique		94	
txa3877	33.71	unique		94	
umc16	-	region			
txa6204	35.05	unique		94	
txa2984	**32.08**	**framework**	**135.6**	**94**	
txa2064	39.33	unique		94	
txa383	36.23	unique			
txa2179	24.62	unique		94	
txa6120	28.18	unique		94	
txp437	30.27	unique		94	CD229025
txa2500	18.31	unique			
txa4107	28.82	region			
txa3520	29.69	unique		94	
txa2982	28.46	region		94	
txa235	**20.87**	**framework**	**140.0**	**94**	
Corn031	35.34	region		94	
txa3518	33.42	unique		94	
TS196	36.03	region		94	
txa428	31.04	unique		94	
txa2243	34.45	region		94	
txa3529	29.88	unique		94	
txa2999	27.4	region		94	
txa2660	19.21	region		94	
txa4068	29.32	region		94	
txa2961	**26.01**	**framework**	**142.5**	**94**	
txa2877	**30.85**	**framework**	**144.4**		
txs422	**31.66**	**framework**	**145.5**		
txs1927	**38.13**	**framework**	**145.5**		
txa2111	28.17	region		26	
txa443	30.18	region		26	
BH245191	34.75	unique		26	SHO46
txp438	31.43	unique		26	BG240233
txp439	33.71	unique		26	BG411222
txp440	35.05	region		26	BG463174
txa6093	23.86	region			
txp542	36.33	unique		26	CW492781
txa3409	30.19	region		26	
bnl15.20	**29.03**	**framework**	**147.3**		
umc7.2	28.53	unique			
txp548	29.69	region			18606.Contig2
txa3598	29.6	region			
bcd738	**21.73**	**framework**	**149.1**		
txa4063	**28.17**	**framework**	**149.1**	**26**	
txa6097	31.14	region		26	
txa3789	25.59	region		26	
txp441	35.05	unique		26	BG649632
txa6125	35.05	unique		26	
txa379	32.24	unique		26	

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa3676	36.93	unique			
txa2986	**32.59**	**framework**	**150.3**		
SDB009	35.73	unique		26	
txa6319	30.46	region			
txa2937	**28.17**	**framework**	**151.9**	**26**	
txp446	32.7	unique		26	7427.Contig1
txp442	27.6	unique		26	BM331329
txa3242	27.89	region		26	
txa6351	31.04	region		26	
isu166	**-**	**framework**	**156.2**		
txa3331	35.05	region		26	
txa6040	31.66	unique		26	
txp447	33	unique		26	5978.Contig2
txa3899	34	region		26	
txs1226	32.54	unique			
txa2271	**22.77**	**framework**	**158.3**	**26**	
txa6296	29.6	region		26	
txp38	25.86	unique		26	
txp421	34.15	unique		26	CW091071
txp449	32.1	region		26	CD231326
txa2515	32.24	unique		26	
txp448	33	unique			23242.Contig1
txa3390	35.64	unique			
txa253	22.23	region			
txa2027	**20.76**	**framework**	**161.4**	**26**	
txp422	35.05	region		26	BE360581
txp420	30.85	unique		26	CW035024
txa6283	**23.58**	**framework**	**165.1**		
txa2051	**28.74**	**framework**	**166.9**		
txa2275	27.24	region			
txp34	25.08	unique		26	
txa2830	31.61	region		26	
txa3448	**28.18**	**framework**	**169.0**	**26**	
txa2032	37.83	region		26	
txp424	38.43	region		26	
txa2065	28.52	unique			
txa6043	19.4	unique		26	
txa6041	31.04	unique		26	
txa3108	32.3	unique			
txa2667	22.5	unique		26	
txi11	32.08	region		26	
txa3140	33.42	region			
isu68	**-**	**framework**	**174.5**		
txi13	35.94	unique		26	
isu52.2	31.95	unique		26	
txa3719	26.42	region		26	
txi7	35.73	region		26	
txi12	35.94	unique		26	
txa3891	**23.39**	**framework**	**177.4**	**26**	
txi9	35.64	unique		26	
txa3610	**27.4**	**framework**	**179.8**		

Table X. Continued

Marker Name	LOD[a]	Mapping[b]	cM	HICF Ctg	Notes[c]
txa4157	26.69	region		26	
txa3987	31.66	unique		26	
txa3402	33.42	region			
txa2056	**24.67**	**framework**	**183.0**	**26**	
txa3904	**31.36**	**framework**	**183.8**	**26**	
txa4073	**31.95**	**framework**	**184.6**	**26**	
txa3008	23.58	unique		26	
txa422	32.01	unique			
txp69	21.93	unique		26	
txp427	26.57	unique		26	BG047875
txa2232	32.01	unique			
txa3559	24.4	region		26	
txa3381	30.46	region			
txa341	**18.41**	**framework**	**190.0**	**26**	
CUa1	32.54	region			
TS076	35.34	region		26	
txa245	29.03	region		26	
txp426	35.34	unique		26	
txa3859	**24.9**	**framework**	**192.5**		
txp425	29.32	unique		26	
txa207	26.03	unique		26	
isu52.1	**16.71**	**framework**	**198.5**	**26**	
txi8	29.69	unique		26	
txi10	35.05	region		26	
txa4138	**22.2**	**framework**	**202.4**	**26**	
txi6	36.03	unique		26	
txi5	**35.43**	**framework**	**202.4**	**26**	
txa3511	23.78	off-end			
txa3280	23.8	off-end		26	
txa3801	10.24	off-end			
txa4158	14.29	off-end			

[a]LOD score at which the given marker was assigned to chromosome 3 using Mapmaker software. A – indicates those markers that were chosen as anchors for chromosome 3 and therefore were not assigned a LOD score (Menz et al., 2002).

[b]Markers indicated in bold with a 'framework' designation are those that were chosen to represent the chromosome 3 framework linkage group as described by Menz et al. (2002). All other markers were placed relative to the framework markers and if their placement was in a unique location between two framework markers they are denoted as unique. Those markers that could be placed in more than one region are denoted as region and are placed in the region showing the highest LOD score. Markers that could only be placed off the end of the chromosome are denoted as off-end.

[c]Sorghum EST, GSS and cDNA probes sequences found to contain microsatellite repeats using the program SSRIT (www.gramene.org) were tested for polymorphism in the BTx623 x IS3620C mapping population. Those that were polymorphic were mapped and given a txp designation consistent with all other microsatellite markers developed at Texas A&M University. The GenBank accession number of the corresponding EST in which the SSR resides is indicated. Those GSS sequences that were polymorphic show the contig information for the given sequence and the cDNA probes are denoted by their name from the map of (Bowers et al., 2003).

Table XI. Macrocolinearity analysis between sorghum chromosome 3 and rice chromosome 1 based on mapped EST-STS markers

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
457	CD206094			CD206094		
	BM323096			BM323096		
				CD227587	8642	Lg03
	CD232230			CD232230		
	AW680993			AW680993		
	CB924743			CB924743		
	CF430667			CF430667		
	BG465556			BG465556		
				CB924715	981	Lg01
	CD210538			CD210538		
				AW679881	416	Lg01
				BE361282	10649	
	CD432685			CD432685		
	BE362724			BE362724		
	AW745363			AW745363		
	BG463900			BG463900		
	BE918563			BE918563		
	BG050175			BG050175		
	CD233176			CD233176		
	BM325718			BM325718		
	CF487139			CF487139		
	AW284327			AW284327		
	AW745220			AW745220		
	CB925925			CB925925		
8642	BE360416			BE360416		
	BE600044	77	Lg03			
	BF588212			BF588212		
	BG157951			BG157951		
	BI211952			BI211952		
	CB925111			CB925111		
	CD222077			CD222077		
	CF433750			CF433750		
	BF656758			BF656758		
	BE917960			BE917960		
	BE366333			BE366333		
	BE919246			BE919246		
	CN147465			CN147465		
	CN139306			CN139306		
	BG464249			BG464249		
	BG357137			BG357137		
	CF434298			CF434298		
	BG463007			BG463007		
	BE594891			BE594891		
	BG322410			BG322410		
				BG560609	4	

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	BI098939			BI098939		
				AW565483	9323	
	BG559311			BG559311		
	CN124600			CN124600		
				CF759206	346	Lg01
	CD208444			CD208444		
				BG356158	6054	Lg09
	CD232744			CD232744		
	CF430460			CF430460		
	CD227587	457	Lg03			
	BE361817			BE361817		
	BG558381	Pericentromeric heterochromatic block	Lg03			
	CF073768			CF073768		
	BM325747			BM325747		
	CF432921			CF432921		
	CD221887			CD221887		
498	BM324072			BM324072		
	CF430379			CF430379		
	BE592946			BE592946		
				BG051020	6836	Lg01
	CN125684			CN125684		
30	BE362721			BE362721		
	CN136604			CN136604		
				CF759146	9190	Lg07
23	BM318170			BM318170		
	BG102659			BG102659		
	BE357974			BE357974		
	CN145587			CN145587		
	CF756227			CF756227		
	BI140817			BI140817		
	CN125678			CN125678		
	CN124567			CN124567		
	BE367090			BE367090		
	AW747586			AW747586		
	BE601354			BE601354		
	CD424673			CD424673		
	BG356249			BG356249		
	CD426657			CD426657		
	CD207483			CD207483		
	BM330250			BM330250		
516	CD426024			CD426024		
	BI351046			BI351046		
	CN141041			CN141041		
	BG356366			BG356366		
	BI076023			BI076023		
	CF671597			CF671597		
	BG047931			BG047931		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	CN143494			CN143494		
	AW286991			AW286991		
	CN149581			CN149581		
	BI099513			BI099513		
	CN142783			CN142783		
	CD222337	Pericentromeric heterochromatic block	Lg03			
	CF429665			CF429665		
	BG465511			BG465511		
				AW672183	20	Lg08
	CN124182			CN124182		
9539	CN152712			CN152712		
	CF485686			CF485686		
	CN141545			CN141545		
				CB926100	9519	
	BG355531			BG355531		
8664	BG050925			BG050925		
	CB928400			CB928400		
	CF481175			CF481175		
	BM322298			BM322298		
1232	CN130514			CN130514		
				CN124717	1234	Lg02
	CN128307			CN128307		
	BE357593			BE357593		
	BG606033			BG606033		
	CF759634			CF759634		
	BG464581			BG464581		
	CN134373			CN134373		
	CD232920			CD232920		
	BE361097			BE361097		
	CD209784			CD209784		
1739	AW925208			AW925208		
	BE362031			BE362031		
	CB924950			CB924950		
	CF488263			CF488263		
	BE360523			BE360523		
	BG948694			BG948694		
	BE357022			BE357022		
	CF484711			CF484711		
				CD461699	641	Lg10
	CN143182			CN143182		
	CB926484			CB926484		
	BG357858			BG357858		
	BG412603			BG412603		
	BE360486			BE360486		
	CN147872			CN147872		
	CN137353			CN137353		
	BG412878			BG412878		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	CW404670			CW404670		
	BG241010			BG241010		
	CF755760			CF755760		
				CF489596	268	Lg06
	BG355967			BG355967		
608	BF586278			BF586278		
	BE356086			BE356086		
	CD236246			CD236246		
				CD235256	6317	Lg07
508	BE362021			BE362021		
	BM329651			BM329651		
	CF487919			CF487919		
	CN133314			CN133314		
	CD205523			CD205523		
	BQ656148			BQ656148		
	CD212159	Duplicate Ctg26	Lg03			
	CN146249					
	BG556954			BG556954		
				CN147781	5948	Lg01
77	CF429479			CF429479		
	CF760171			CF760171		
	CD429994			BM322273		
				AW671572	1759	Lg09
	CD212473			BI074314		
	BG357268			BE599043		
	BI245426			BI245426		
	BE599043			BG357268		
	BM322273			CD212473		
	BI074314			CD429994		
	CF071605			CF071605		
	BM326558			BM326558		
	BM324386			BM324386		
	CD213887			CD213887		
				BG103151	176	
				CF433857	1404	Lg07
	CF488778			CF488778		
	AW671612			AW671612		
	CF432097			CF432097		
	CN149581	Duplicate Ctg516	Lg03			
				BF585641	5828	Lg09
	CN138086			CN138086		
	BE356757			BE356757		
	CB925191			CB925191		
	BE595998			BE595998		
	BG241188			BG241188		
	BG357572			BG357572		
	BG273434			BG273434		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	CD424000			CD424000		
	BE918782			BE918782		
	BG322645			BG322645		
	CN139670			CN139670		
	CF489237			CF489237		
	CF770005			CF770005		
	BM330754			BM330754		
	BG947115			BG947115		
	BF481741			BF481741		
	BM323630			BM323630		
				BE600044	8642	Lg03
	BE361859			BE361859		
	BF317919			BF317919		
	CD427864			CD427864		
	CF757987			CF757987		
				BQ656102	106	Lg05
	BG102636			BG102636		
385	CF759039			CF759039		
	BI076054			BI076054		
	BG463285			BG463285		
				CD209645	288	Lg09
	BM322493			BM322493		
	BI074369			BI074369		
	CF430703			CF430703		
	CF483636			CF483636		
	CD432352			CD432352		
	BG484761			BG484761		
	CF486194			CF486194		
	BF587445			BF587445		
	BF705046			BF705046		
	CD430678			CD430678		
	BG050591			BG050591		
	BF422147			BF422147		
	AW564197			AW564197		
	BE917681			BE917681		
	AW564023			AW564023		
				CD223194	1759	Lg09
	BI074560			BI074560		
	BG048220			BG048220		
	CF427926			CF427926		
	CD232508			CD232508		
	CD432311			CD432311		
	BG557844			BG557844		
	CF487081			CF487081		
	BG557733			BG557733		
	BE364599			BE364599		
	BI643733			BI643733		
	CD207141			CD207141		
	BE355272			BE355272		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
1560	CN128851			CN128851		
				CD228613	456	Lg01
	BI211911			BI211911		
				CD426663	1208	Lg01
	BE125240			BE125240		
	CD423226			CD423226		
4076	BQ656056			BQ656056		
	BG556314			BG556314		
	BF585483			BF585483		
	CD236122			CD236122		
	AW283704			AW283704		
	BE595281			BE595281		
	BG605885			BG605885		
	BI074822			BI074822		
	CD219775			CD219775		
	CN131469			CN131469		
4611	CD206443			CD206443		
	AW745905			AW745905		
	CF489789			CF489789		
	CD423600			CD423600		
	CN139135			CN139135		
	BG947283			BG947283		
	BG605849			BG605849		
	BG487823			BG487823		
	CF433257			CF433257		
	BG412082			BG412082		
	CD210345			CD210345		
94	BG464084			BG464084		
	CD208671			CD208671		
	BE362379			BE362379		
	BF587140			BF587140		
	BM322554			BM322554		
				BF586254	725	
	CF432315			CF432315		
	BE599372			BE599372		
	BG356353			BG356353		
				AW680206	6477	
	BM324066			BM324066		
	CD430563			CD430563		
	CF433657			CF433657		
	BE598972			BE598972		
	AW671043			AW671043		
	BF585523			BF585523		
	BE597345			BE597345		
	BE593007			BE593007		
	BG322482			BG322482		
	BF588141			BF588141		
	AW286905			AW286905		
				BG464054	161	Lg09

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	CN128558	Pericentromeric heterochromatic block	Lg03			
	CD219796			CD219796		
	CD427711			CD427711		
	CD223691			CD223691		
	BF422011			BF422011		
	CF489408			CF489408		
	BE355052			BE355052		
	BE592309			BE592309		
	AW678813			AW678813		
	CF072785			CF072785		
	BG102297			BG102297		
	CD209633			CD209633		
	BF176974			BF176974		
	AW746273			AW746273		
	BI076163			BI076163		
	CF071759			CF071759		
	BF587193			BF587193		
	BF656208			BF656208		
	CN125212			CN125212		
26	AW922570			AW922570		
	BI141001			BI141001		
	CD462481			CD462481		
	CD428230			CD428230		
	CF483479			CF483479		
	CD463905			CD463905		
	CN139743	Pericentromeric heterochromatic block	Lg03			
	BE357360			BE357360		
	CD232661			CD232661		
	CF489057			CF489057		
				BG649183	Singletons	
	BI351257			BI351257		
	BM323001			BM323001		
	CD426427			CD426427		
	BI140669			BI140669		
	BG411084			BG411084		
	BG739642			BG739642		
	AW679507			AW679507		
	BE594341			BE594341		
	CD221368			CD221368		
	BM325547			BM325547		
	BF704882			BF704882		
	BF657709			BF657709		
	CD428050			CD428050		
	BF588344			BF588344		
	BE918679			BE918679		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	BI245429			BI245429		
	CD232634			CD232634		
	CB925609			CB925609		
	CD207824			CD207824		
	BE366178			BE366178		
	BE918964			BE918964		
	BM323409			BM323409		
	CD432487			CD432487		
	CD233877			CD233877		
	CD212159			CD212159		
	BM323382			BM323382		
	CD205497			CD205497		
	BF587198			BF587198		
	CB926739			CB926739		
	CD211911			CD211911		
	BI099158			BI099158		
				CB929314	456	Lg07
	CN129546			CN129546		
	CD424655			CD424655		
	CD230557			CD230557		
	BE598736			BE598736		
	BE595800			BE595800		
	AW745704			AW745704		
	BE599902			BE599902		
	CD204322			CD204322		
	BG559010			BG559010		
				BG101892	6006	Lg04
	BG557643			BG557643		
	CN135824			CN135824		
	BE363696			BE363696		
	BF656081			BF656081		
	BE592353			BE592353		
	BF587256			BF587256		
	AW565362			AW565362		
	BE599995			BE599995		
				BG462935	164	Lg09
	BG948452			BG948452		
	BG462899			BG462899		
	AW286613			AW286613		
	BG559251			BG559251		
	BM318017			BM318017		
	BF656749			BF656749		
	BM324028			BM324028		
	BI211961			BI211961		
	BE600286			BE600286		
	BE125224			BE125224		
	BE367030	Pericentromeric heterochromatic block	Lg03			

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	BE355074			BE355074		
	BM323907			BM323907		
	BM323700			BM323700		
	BE600323			BE600323		
	BG947252			BG947252		
	BE600064			BE600064		
	BG488122			BG488122		
	BG050777			BG050777		
	BM324945			BM324945		
	BE358182			BE358182		
	BM325245			BE598591		
	BG101800			BM325245		
	BE598591			BG101800		
	BG817523			BG817523		
	BE358320			BE358320		
	AW283346			AW283346		
	BG946972			BG946972		
	AW745953			AW745953		
	AW679713			AW679713		
	BG356230			BG356230		
	BG356592			BG356592		
	BM322880			BM322880		
	BM324063			BM324063		
	BG465378			BG465378		
	BE918371			BE918371		
	AW745139			AW745139		
	BF656904			BF656904		
	BG159525			BG159525		
	BM325684			BM325684		
	BE361743			BE361743		
	BG649801			BG649801		
	AW678633			AW678633		
	BM323894			BM323894		
	BG947288			BG947288		
	BG357433			BG357433		
	BG410959			BG410959		
	BG322470			BG322470		
	BF705560			BF705560		
				AW563509	26	Lg03
	BE597021			BE597021		
				BI098434	7726	Lg10
	BF481723			BF481723		
	BE597990			BE597990		
	BF585506			BF585506		
	BE360148			BE360148		
	BG051583			BG051583		
	AW563509	26	Lg03			
	BG159264			BG159264		
	BE597471			BE597471		

Table XI. Continued

Sorghum ESTs on sorghum chromosome 3 minimal tiling path[a]				Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path[b]		
Contig	EST	Move from contig	Linkage group	EST	Move to contig	Linkage group
	BE357331			BE357331		
	BI075560			BI075560		
	BE597676			BE597676		
	BE360847			BE360847		
	BG932945			BG932945		
	AW679896			AW679896		
				BE595807	3654	
	BG556494			BG556494		
	BE917877			BE917877		
	AW564879			AW564879		
	BM325870			BM325870		
				BI075954	275	Lg08
	BE594020			BE594020		
				BF421983	1102	
				BI351450	1102	
				BM323644	237	Lg01
	AW924385			AW924385		
	BG158627			BG158627		
	BG487560			BG487560		
	BG560112			BG560112		
	BE358709			BE358709		
	AW924587			AW924587		
	BE593646			BE593646		
				BI211591	6317	Lg07
	BI211668			BI211668		
	BG948567			BG948567		
				BE598794	5828	Lg09
				BM323684	471	Lg01
				BE597304	5828	Lg09
				BI139681	183	Lg05
				BG101724	183	Lg05
				BM325592	10398	

[a]Sorghum ESTs mapped on the sorghum chromosome 3 minimal tiling path are listed. Contigs are listed in the order in which they align to the genetic map. If an EST is annotated with 'move from contig' and the corresponding 'Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path' are empty, it indicates the move in of this EST into its current location relative to the rice chromosome 1 minimal tiling path. Linkage group information is listed if available.

[b]Sorghum ESTs homeologous to rice chromosome 1 minimal tiling path are listed. If an EST is annotated with 'move to contig' and the corresponding 'Sorghum ESTs mapped on sorghum chromosome 3 minimal tiling path' are empty, it indicates the move out of this EST relative to the sorghum minimal tiling path of chromosome 3. Linkage group information is listed if available.

Table XII. BLASTX analysis of sorghum pool skim sequences to the TIGR non-TE rice gene models (TIGR Release 3)

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 1	Pool1.Contig159	LOC_Os01g60600.1	2e-33	144	100/221
Pool 1	Pool1.Contig131	LOC_Os01g60640.1	2e-49	123	86/185
Pool 1	Pool1.Contig27	LOC_Os01g60650.1	7e-43	139	72/94
Pool 1	Pool1.Contig175	LOC_Os01g60660.1	1e-41	110	55/76
Pool 1	Pool1.Contig120	LOC_Os01g60670.1	e-127	454	249/318
Pool 1	Pool1.Contig140	LOC_Os01g60670.1	0.0	652	327/354
Pool 1	Pool1.Contig114	LOC_Os01g60700.2	4e-11	69	36/45
Pool 1	Pool1.Contig171	LOC_Os01g60700.2	7e-95	348	178/206
Pool 1	Pool1.Contig42	LOC_Os01g60730.1	8e-43	172	96/138
Pool 1	Pool1.Contig137	LOC_Os01g60770.1	2e-52	205	112/172
Pool 1	Pool1.Contig88	LOC_Os01g60800.1	7e-27	120	54/77
Pool 1	it55d05_00.g1.b	LOC_Os01g60810.1	1e-28	122	74/122
Pool 1	Pool1.Contig49	LOC_Os01g60910.1	1e-47	189	98/147
Pool 1	Pool1.Contig96	LOC_Os01g60910.1	e-143	508	255/294
Pool 1	Pool1.Contig143	LOC_Os01g60930.1	2e-87	222	112/162
Pool 1	Pool1.Contig156	LOC_Os01g60930.1	2e-21	103	52/64
Pool 1	Pool1.Contig13	LOC_Os01g60940.1	2e-21	100	63/143
Pool 1	Pool1.Contig154	LOC_Os01g60960.1	2e-83	167	96/147
Pool 1	it56f05_00.g1.b	LOC_Os01g61010.1	3e-66	248	116/127
Pool 1	Pool1.Contig75	LOC_Os01g61010.1	1e-47	147	74/106
Pool 1	qt28f12_00.b1.b	LOC_Os01g61020.1	4e-15	78	51/128
Pool 1	Pool1.Contig60	LOC_Os01g61020.1	4e-21	100	69/148
Pool 1	Pool1.Contig147	LOC_Os01g61020.1	3e-33	102	60/123
Pool 1	Pool1.Contig19	LOC_Os01g61030.1	3e-23	105	56/90
Pool 1	Pool1.Contig178	LOC_Os01g61030.1	e-104	310	179/301
Pool 1	Pool1.Contig167	LOC_Os01g61044.1	e-115	307	170/214
Pool 1	Pool1.Contig173	LOC_Os01g61044.2	2e-72	201	110/140
Pool 1	Pool1.Contig63	LOC_Os01g61060.1	e-161	466	222/308
Pool 1	Pool1.Contig41	LOC_Os01g61080.1	6e-45	179	107/187
Pool 1	Pool1.Contig89	LOC_Os01g61110.1	5e-56	217	113/161
Pool 1	Pool1.Contig65	LOC_Os01g61120.1	7e-17	87	59/106
Pool 1	Pool1.Contig80	LOC_Os01g61120.1	8e-16	83	37/50
Pool 1	Pool1.Contig134	LOC_Os01g61120.1	e-108	335	163/188
Pool 1	Pool1.Contig32	LOC_Os01g61160.1	3e-63	239	107/118
Pool 1	Pool1.Contig90	LOC_Os01g61160.1	2e-82	282	144/183
Pool 1	qt28f07_00.b1.b	LOC_Os01g61180.1	1e-62	234	113/120
Pool 1	Pool1.Contig09	LOC_Os01g61180.1	4e-71	264	125/153
Pool 1	Pool1.Contig138	LOC_Os01g61190.1	e-153	320	160/210
Pool 1	Pool1.Contig28	LOC_Os01g61210.1	5e-15	74	38/63
Pool 1	Pool1.Contig95	LOC_Os01g61230.1	2e-74	151	75/137
Pool 1	Pool1.Contig160	LOC_Os01g61230.1	4e-55	167	84/150
Pool 1	Pool1.Contig34	LOC_Os01g61250.1	1e-34	144	68/88
Pool 1	Pool1.Contig58	LOC_Os01g61250.1	3e-42	171	88/164
Pool 1	Pool1.Contig135	LOC_Os01g61250.1	9e-46	184	96/157
Pool 1	Pool1.Contig22	LOC_Os01g61310.1	9e-70	261	127/160
Pool 1	Pool1.Contig106	LOC_Os01g61320.1	3e-74	278	151/235
Pool 1	it61g06_00.g1.b	LOC_Os01g61460.1	2e-62	236	104/120
Pool 1	Pool1.Contig165	LOC_Os01g61460.1	4e-50	199	119/245

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 1	Pool1.Contig79	LOC_Os01g61470.1	3e-22	104	65/120
Pool 1	Pool1.Contig174	LOC_Os01g61480.1	5e-31	136	94/221
Pool 1	Pool1.Contig25	LOC_Os01g61500.1	4e-49	193	104/148
Pool 1	Pool1.Contig72	LOC_Os01g61500.1	1e-20	99	47/62
Pool 1	Pool1.Contig107	LOC_Os01g61500.1	8e-16	84	36/40
Pool 1	Pool1.Contig169	LOC_Os01g61550.1	0.0	430	203/218
Pool 1	Pool1.Contig37	LOC_Os01g61580.1	3e-25	110	50/53
Pool 1	Pool1.Contig125	LOC_Os01g61580.1	5e-20	66	39/64
Pool 1	Pool1.Contig108	LOC_Os01g61590.1	4e-48	191	101/131
Pool 1	Pool1.Contig146	LOC_Os01g61590.1	1e-49	109	57/69
Pool 1	Pool1.Contig179	LOC_Os01g61610.4	2e-65	251	128/192
Pool 1	Pool1.Contig94	LOC_Os01g61620.1	9e-49	192	93/100
Pool 1	Pool1.Contig61	LOC_Os01g61630.1	1e-52	204	114/164
Pool 1	Pool1.Contig132	LOC_Os01g61630.1	0.0	465	240/307
Pool 1	Pool1.Contig71	LOC_Os01g61640.1	4e-34	144	80/123
Pool 1	Pool1.Contig109	LOC_Os01g61640.1	4e-11	68	32/35
Pool 1	Pool1.Contig119	LOC_Os01g61670.1	5e-39	160	97/213
Pool 1	Pool1.Contig83	LOC_Os01g61680.1	2e-47	189	104/151
Pool 1	Pool1.Contig35	LOC_Os02g49670.1	2e-11	68	78/273
Pool 1	Pool1.Contig177	LOC_Os03g57100.1	6e-26	120	64/100
Pool 1	Pool1.Contig110	LOC_Os04g34100.1	2e-46	172	84/103
Pool 1	Pool1.Contig121	LOC_Os05g39080.1	4e-46	120	62/110
Pool 1	Pool1.Contig21	LOC_Os05g39670.1	6e-85	256	132/169
Pool 1	Pool1.Contig10	LOC_Os05g39860.1	6e-15	79	68/227
Pool 1	Pool1.Contig126	LOC_Os05g39860.1	4e-15	74	40/91
Pool 1	Pool1.Contig163	LOC_Os08g35470.1	9e-51	201	115/162
Pool 1	Pool1.Contig68	LOC_Os08g42750.1	1e-40	100	45/55
Pool 1	Pool1.Contig164	LOC_Os08g44480.2	3e-29	96	47/57
Pool 1	Pool1.Contig172	LOC_Os11g42640.1	9e-87	213	104/211
Pool 1	Pool1.Contig118	LOC_Os12g20370.1	9e-11	51	43/162
Pool 1	Pool1.Contig67	LOC_Os12g35570.1	7e-68	170	86/132
Pool 2	Pool2.Contig172	LOC_Os01g01820.1	3e-11	68	30/66
Pool 2	Pool2.Contig219	LOC_Os01g21330.1	1e-48	194	110/194
Pool 2	ix04a04.b1.b	LOC_Os01g43390.1	2e-47	187	100/247
Pool 2	Pool2.Contig36	LOC_Os01g61590.1	7e-40	107	54/57
Pool 2	Pool2.Contig113	LOC_Os01g61590.1	6e-20	97	46/47
Pool 2	ix07d06.b1.b	LOC_Os01g61610.3	2e-12	68	30/49
Pool 2	Pool2.Contig82	LOC_Os01g61610.4	1e-32	139	66/79
Pool 2	ix01c03.b1.b	LOC_Os01g61620.1	3e-41	100	46/61
Pool 2	ix01f12.g1.b	LOC_Os01g61620.1	3e-14	77	43/56
Pool 2	Pool2.Contig215	LOC_Os01g61620.1	2e-75	192	93/100
Pool 2	Pool2.Contig96	LOC_Os01g61630.1	e-119	427	216/263
Pool 2	Pool2.Contig45	LOC_Os01g61640.1	2e-14	78	38/38
Pool 2	Pool2.Contig128	LOC_Os01g61640.1	4e-31	104	47/54
Pool 2	Pool2.Contig158	LOC_Os01g61670.2	3e-51	200	93/117
Pool 2	Pool2.Contig145	LOC_Os01g61680.1	2e-47	189	104/151
Pool 2	Pool2.Contig202	LOC_Os01g61690.1	7e-70	173	77/87
Pool 2	Pool2.Contig176	LOC_Os01g61720.6	3e-51	201	117/185
Pool 2	Pool2.Contig169	LOC_Os01g61720.7	2e-77	289	149/188
Pool 2	Pool2.Contig47	LOC_Os01g61780.1	2e-14	79	36/42
Pool 2	Pool2.Contig49	LOC_Os01g61780.1	5e-42	169	94/157

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 2	Pool2.Contig155	LOC_Os01g61780.1	3e-29	128	59/67
Pool 2	Pool2.Contig195	LOC_Os01g61780.1	2e-53	148	84/137
Pool 2	Pool2.Contig72	LOC_Os01g61820.1	4e-66	179	92/123
Pool 2	Pool2.Contig139	LOC_Os01g61830.1	7e-23	106	53/61
Pool 2	Pool2.Contig194	LOC_Os01g61830.1	2e-32	139	63/66
Pool 2	Pool2.Contig171	LOC_Os01g61860.1	3e-19	95	54/91
Pool 2	Pool2.Contig229	LOC_Os01g61860.1	e-111	305	172/266
Pool 2	ix03b07.b1.b	LOC_Os01g61880.1	4e-17	86	36/42
Pool 2	ix05d02.b1.b	LOC_Os01g61880.1	7e-62	106	47/50
Pool 2	Pool2.Contig214	LOC_Os01g61880.2	7e-86	145	68/83
Pool 2	Pool2.Contig101	LOC_Os01g61890.4	2e-43	174	81/125
Pool 2	Pool2.Contig226	LOC_Os01g61900.1	1e-34	147	83/148
Pool 2	ix05f05.g1.b	LOC_Os01g61910.1	1e-56	216	125/210
Pool 2	Pool2.Contig100	LOC_Os01g61910.1	e-129	445	226/315
Pool 2	Pool2.Contig223	LOC_Os01g61910.1	2e-76	286	146/174
Pool 2	Pool2.Contig211	LOC_Os01g61930.1	3e-38	158	88/170
Pool 2	Pool2.Contig220	LOC_Os01g61930.1	2e-86	279	150/247
Pool 2	ix03a04.g1.b	LOC_Os01g61930.2	8e-49	191	97/160
Pool 2	Pool2.Contig161	LOC_Os01g61940.1	e-145	512	287/371
Pool 2	Pool2.Contig190	LOC_Os01g61940.1	7e-40	163	78/99
Pool 2	Pool2.Contig210	LOC_Os01g61970.1	4e-34	144	73/94
Pool 2	Pool2.Contig212	LOC_Os01g61970.1	e-175	612	299/321
Pool 2	Pool2.Contig224	LOC_Os01g61980.1	e-171	602	307/361
Pool 2	Pool2.Contig228	LOC_Os01g61990.1	e-143	509	258/278
Pool 2	Pool2.Contig208	LOC_Os01g62000.1	4e-70	245	114/146
Pool 2	Pool2.Contig160	LOC_Os01g62010.1	5e-31	112	51/76
Pool 2	Pool2.Contig61	LOC_Os01g62030.1	1e-49	195	108/164
Pool 2	Pool2.Contig173	LOC_Os01g62040.2	2e-24	112	57/63
Pool 2	Pool2.Contig78	LOC_Os01g62060.1	7e-12	70	31/35
Pool 2	Pool2.Contig125	LOC_Os01g62070.1	1e-35	147	77/112
Pool 2	Pool2.Contig163	LOC_Os01g62070.1	5e-15	80	37/39
Pool 2	Pool2.Contig22	LOC_Os01g62080.1	2e-30	131	57/59
Pool 2	Pool2.Contig130	LOC_Os01g62080.1	3e-31	134	72/120
Pool 2	Pool2.Contig153	LOC_Os01g62080.1	6e-36	149	77/118
Pool 2	Pool2.Contig156	LOC_Os01g62080.1	1e-15	84	55/114
Pool 2	Pool2.Contig196	LOC_Os01g62110.1	4e-29	127	74/91
Pool 2	Pool2.Contig147	LOC_Os01g62160.1	e-172	602	287/355
Pool 2	Pool2.Contig216	LOC_Os01g62190.1	1e-55	216	128/200
Pool 2	ix02c06.g1.b	LOC_Os01g62200.1	2e-44	176	101/149
Pool 2	Pool2.Contig63	LOC_Os01g62200.1	1e-24	108	60/100
Pool 2	Pool2.Contig112	LOC_Os01g62220.1	e-117	418	203/239
Pool 2	Pool2.Contig193	LOC_Os01g62230.1	2e-66	251	131/139
Pool 2	ix03f06.g1.b	LOC_Os01g62260.1	3e-25	113	57/69
Pool 2	Pool2.Contig104	LOC_Os01g62310.1	2e-15	82	71/152
Pool 2	Pool2.Contig142	LOC_Os01g62310.1	2e-46	185	105/163
Pool 2	Pool2.Contig86	LOC_Os01g62390.1	1e-50	117	52/58
Pool 2	Pool2.Contig181	LOC_Os01g62410.2	2e-93	341	196/333
Pool 2	Pool2.Contig222	LOC_Os01g62410.2	2e-25	117	65/133
Pool 2	Pool2.Contig52	LOC_Os01g62420.1	2e-22	104	69/158
Pool 2	Pool2.Contig178	LOC_Os01g62420.4	6e-15	81	39/44
Pool 2	ix06d01.g1.b	LOC_Os01g62440.1	2e-29	126	60/64

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 2	ix07a03.b1.b	LOC_Os01g62440.1	4e-35	145	69/78
Pool 2	Pool2.Contig143	LOC_Os01g62440.1	5e-60	123	60/81
Pool 2	Pool2.Contig151	LOC_Os01g62440.1	7e-64	182	104/185
Pool 2	ix03d06.g1.b	LOC_Os01g62460.2	1e-83	306	163/257
Pool 2	ix01f07.g1.b	LOC_Os01g62490.1	1e-21	101	57/97
Pool 2	Pool2.Contig70	LOC_Os01g62490.1	4e-93	339	170/196
Pool 2	ix03d02.b1.b	LOC_Os01g62500.1	1e-73	235	118/144
Pool 2	Pool2.Contig20	LOC_Os01g62500.1	3e-20	86	62/123
Pool 2	ix04h03.b1.b	LOC_Os01g62584.1	3e-41	166	96/175
Pool 2	Pool2.Contig29	LOC_Os01g62620.1	5e-21	79	33/36
Pool 2	Pool2.Contig120	LOC_Os01g62620.1	7e-14	77	34/57
Pool 2	ix05c06.g1.b	LOC_Os01g62630.1	6e-39	158	71/79
Pool 2	ix03c05.b1.b	LOC_Os01g62650.1	6e-27	119	59/60
Pool 2	Pool2.Contig111	LOC_Os01g62650.1	5e-83	306	162/209
Pool 2	ix02c07.b1.b	LOC_Os01g62760.1	5e-54	155	77/110
Pool 2	ix06g02.b1.b	LOC_Os01g62790.1	6e-39	137	67/117
Pool 2	ix07e04.b1.b	LOC_Os01g62790.1	6e-18	88	39/50
Pool 2	Pool2.Contig75	LOC_Os01g62800.1	1e-39	161	74/100
Pool 2	Pool2.Contig39	LOC_Os02g06650.1	3e-13	75	37/60
Pool 2	Pool2.Contig28	LOC_Os05g22840.1	8e-23	105	51/54
Pool 2	ix02e02.g1.b	LOC_Os05g38490.1	8e-55	123	56/69
Pool 2	Pool2.Contig54	LOC_Os05g38530.1	3e-74	276	139/145
Pool 2	Pool2.Contig217	LOC_Os05g38550.1	1e-31	97	57/94
Pool 2	Pool2.Contig152	LOC_Os05g38580.1	2e-11	40	19/19
Pool 2	Pool2.Contig159	LOC_Os05g38580.2	2e-16	86	38/50
Pool 2	Pool2.Contig157	LOC_Os05g38820.1	6e-32	122	60/76
Pool 2	ix07f02.b1.b	LOC_Os05g50690.1	2e-69	259	125/152
Pool 2	qt33a12.b1.b	LOC_Os06g24704.1	2e-30	130	67/110
Pool 2	Pool2.Contig76	LOC_Os06g24704.1	4e-31	134	65/82
Pool 2	ix04a04.g1.b	LOC_Os06g45330.1	9e-17	85	47/111
Pool 2	ix01g12.g1.b	LOC_Os06g46620.1	3e-24	110	50/109
Pool 2	Pool2.Contig19	LOC_Os07g08170.1	1e-12	54	25/26
Pool 2	Pool2.Contig175	LOC_Os07g08170.1	3e-19	95	47/51
Pool 2	ix07e08.b1.b	LOC_Os09g20350.1	1e-30	119	61/72
Pool 2	Pool2.Contig218	LOC_Os09g29150.1	1e-15	84	47/79
Pool 2	Pool2.Contig183	LOC_Os09g32360.1	2e-95	348	188/282
Pool 2	Pool2.Contig213	LOC_Os09g32360.1	7e-44	177	96/159
Pool 2	qt33c05.b1.b	LOC_Os10g37660.2	9e-26	114	79/254
Pool 2	ix07c08.g1.b	LOC_Os12g01290.1	2e-19	94	49/123
Pool 2	Pool2.Contig74	LOC_Os12g07540.1	4e-11	67	39/88
Pool 2	Pool2.Contig43	LOC_Os12g08220.1	2e-96	204	109/170
Pool 2	Pool2.Contig189	LOC_Os12g26940.1	2e-30	133	63/119
Pool 2	Pool2.Contig197	LOC_Os12g34030.1	4e-40	120	54/162
Pool 3	ix12a12.g1.b	LOC_Os01g62810.1	9e-41	164	87/134
Pool 3	ix10b12.b1.b	LOC_Os01g62810.2	1e-26	117	56/84
Pool 3	Pool3.Contig97	LOC_Os01g62820.1	3e-16	84	40/57
Pool 3	Pool3.Contig185	LOC_Os01g62820.1	6e-69	262	147/187
Pool 3	Pool3.Contig165	LOC_Os01g62850.1	1e-20	100	53/66
Pool 3	Pool3.Contig184	LOC_Os01g62860.1	3e-80	299	152/217
Pool 3	Pool3.Contig170	LOC_Os01g62870.1	2e-77	163	73/92
Pool 3	Pool3.Contig175	LOC_Os01g62880.2	9e-29	127	61/81

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 3	Pool3.Contig188	LOC_Os01g62900.1	e-107	187	113/193
Pool 3	Pool3.Contig187	LOC_Os01g62910.1	0.0	1112	550/657
Pool 3	Pool3.Contig123	LOC_Os01g62920.1	2e-21	103	46/55
Pool 3	Pool3.Contig164	LOC_Os01g62920.1	5e-73	274	150/206
Pool 3	Pool3.Contig130	LOC_Os01g62950.1	7e-60	166	83/95
Pool 3	Pool3.Contig96	LOC_Os01g62990.1	e-143	303	159/215
Pool 3	Pool3.Contig117	LOC_Os01g63010.1	2e-45	103	58/114
Pool 3	Pool3.Contig27	LOC_Os01g63050.1	3e-53	207	122/234
Pool 3	Pool3.Contig86	LOC_Os01g63060.1	2e-19	92	42/53
Pool 3	Pool3.Contig105	LOC_Os01g63060.1	5e-41	100	51/83
Pool 3	Pool3.Contig189	LOC_Os01g63160.1	9e-93	341	166/247
Pool 3	ix11e12.g1.b	LOC_Os01g63180.1	5e-18	89	40/56
Pool 3	Pool3.Contig56	LOC_Os01g63180.1	1e-63	240	127/174
Pool 3	Pool3.Contig129	LOC_Os01g63180.1	2e-13	75	33/43
Pool 3	Pool3.Contig178	LOC_Os01g63190.1	e-137	436	225/319
Pool 3	Pool3.Contig146	LOC_Os01g63190.2	1e-13	77	32/41
Pool 3	Pool3.Contig134	LOC_Os01g63200.1	0.0	548	271/380
Pool 3	Pool3.Contig101	LOC_Os01g63210.1	7e-96	348	173/222
Pool 3	Pool3.Contig136	LOC_Os01g63220.2	e-125	448	208/234
Pool 3	Pool3.Contig182	LOC_Os01g63230.1	e-159	560	291/433
Pool 3	Pool3.Contig173	LOC_Os01g63240.1	3e-25	115	68/130
Pool 3	Pool3.Contig176	LOC_Os01g63250.1	e-100	364	192/248
Pool 3	ix16e11.b1.b	LOC_Os01g63260.1	e-101	363	179/240
Pool 3	Pool3.Contig128	LOC_Os01g63270.1	1e-56	125	58/64
Pool 3	Pool3.Contig158	LOC_Os01g63270.1	9e-93	210	116/172
Pool 3	Pool3.Contig167	LOC_Os01g63270.1	6e-20	78	34/42
Pool 3	Pool3.Contig169	LOC_Os01g63270.1	9e-40	129	61/62
Pool 3	Pool3.Contig58	LOC_Os01g63280.1	8e-35	94	45/56
Pool 3	Pool3.Contig85	LOC_Os01g63280.1	3e-48	190	98/129
Pool 3	Pool3.Contig153	LOC_Os01g63290.1	5e-84	236	120/171
Pool 3	Pool3.Contig51	LOC_Os01g63310.1	e-108	388	210/274
Pool 3	Pool3.Contig87	LOC_Os01g63354.1	2e-92	338	164/227
Pool 3	Pool3.Contig120	LOC_Os01g63354.1	2e-14	79	37/56
Pool 3	ix14d04.b1.b	LOC_Os01g63400.2	3e-24	107	51/57
Pool 3	Pool3.Contig37	LOC_Os01g63400.2	9e-86	314	151/204
Pool 3	Pool3.Contig126	LOC_Os01g63410.1	1e-66	252	131/159
Pool 3	Pool3.Contig17	LOC_Os01g63420.1	3e-23	107	45/61
Pool 3	Pool3.Contig77	LOC_Os01g63470.1	e-157	551	268/326
Pool 3	Pool3.Contig102	LOC_Os01g63480.1	e-142	504	246/279
Pool 3	Pool3.Contig141	LOC_Os01g63480.1	1e-11	70	40/64
Pool 3	Pool3.Contig118	LOC_Os01g68950.3	1e-14	80	42/90
Pool 3	Pool3.Contig34	LOC_Os02g06830.2	4e-14	76	34/81
Pool 3	Pool3.Contig44	LOC_Os02g11650.1	2e-19	81	53/169
Pool 3	Pool3.Contig94	LOC_Os02g17260.1	4e-13	74	45/133
Pool 3	ix16d05.g1.b	LOC_Os02g27340.1	8e-15	78	46/108
Pool 3	ix11f03.b1.b	LOC_Os02g30630.1	2e-31	133	71/115
Pool 3	ix14f11.g1.b	LOC_Os02g58340.1	1e-12	72	39/113
Pool 3	ix14f11.b1.b	LOC_Os03g06740.1	1e-12	70	33/60
Pool 3	Pool3.Contig106	LOC_Os03g25380.1	1e-28	125	72/119
Pool 3	ix14g12.b1.b	LOC_Os03g31550.2	6e-17	86	64/239
Pool 3	ix11e08.g1.b	LOC_Os03g50510.1	1e-21	101	51/120

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 3	Pool3.Contig125	LOC_Os04g13750.1	2e-91	261	129/283
Pool 3	ix11g10.g1.b	LOC_Os04g16680.1	6e-39	158	74/75
Pool 3	Pool3.Contig111	LOC_Os04g16680.1	2e-42	172	89/95
Pool 3	Pool3.Contig145	LOC_Os04g16680.1	1e-27	122	60/62
Pool 3	Pool3.Contig45	LOC_Os04g35060.1	2e-17	88	53/90
Pool 3	ix14h06.b1.b	LOC_Os05g37900.1	1e-21	100	44/53
Pool 3	Pool3.Contig92	LOC_Os05g38000.1	9e-42	109	52/83
Pool 3	ix14b10.g1.b	LOC_Os06g07860.1	2e-29	127	70/181
Pool 3	qt34h06.b1.b	LOC_Os06g16980.2	5e-24	108	68/186
Pool 3	Pool3.Contig180	LOC_Os06g19260.1	0.0	712	447/967
Pool 3	Pool3.Contig30	LOC_Os07g22560.1	6e-38	155	97/262
Pool 3	Pool3.Contig103	LOC_Os08g05800.1	1e-13	76	38/89
Pool 3	Pool3.Contig90	LOC_Os08g07810.1	e-124	443	216/285
Pool 3	Pool3.Contig84	LOC_Os08g13930.2	1e-24	111	65/104
Pool 3	Pool3.Contig179	LOC_Os08g13930.2	e-162	572	299/377
Pool 3	Pool3.Contig42	LOC_Os09g23810.1	4e-20	97	50/124
Pool 3	ix16c08.b1.b	LOC_Os10g22220.1	2e-17	87	55/143
Pool 3	Pool3.Contig151	LOC_Os10g32760.1	6e-15	54	27/35
Pool 3	Pool3.Contig07	LOC_Os11g03040.1	3e-11	67	45/119
Pool 3	qt34f04.b1.b	LOC_Os12g17910.1	1e-14	77	36/73
Pool 4	ix31d11.g1.b	LOC_Os01g63480.1	7e-78	286	138/154
Pool 4	Pool4.Contig11	LOC_Os01g63480.1	4e-75	279	142/182
Pool 4	Pool4.Contig100	LOC_Os01g63500.1	1e-35	150	86/127
Pool 4	ix31g02.g1.b	LOC_Os01g63510.1	1e-57	220	113/131
Pool 4	Pool4.Contig127	LOC_Os01g63510.1	1e-24	114	62/76
Pool 4	Pool4.Contig122	LOC_Os01g63580.1	e-101	367	186/203
Pool 4	Pool4.Contig131	LOC_Os01g63580.1	7e-23	107	51/54
Pool 4	Pool4.Contig144	LOC_Os01g63580.1	2e-28	127	61/65
Pool 4	Pool4.Contig53	LOC_Os01g63620.1	4e-57	219	120/198
Pool 4	Pool4.Contig33	LOC_Os01g63690.1	6e-88	322	164/188
Pool 4	Pool4.Contig124	LOC_Os01g63710.1	6e-36	74	34/36
Pool 4	Pool4.Contig59	LOC_Os01g63710.2	5e-56	217	134/273
Pool 4	Pool4.Contig84	LOC_Os01g63770.1	9e-55	125	61/72
Pool 4	Pool4.Contig115	LOC_Os01g63770.1	4e-38	157	73/79
Pool 4	Pool4.Contig113	LOC_Os01g63780.1	1e-15	82	38/41
Pool 4	Pool4.Contig135	LOC_Os01g63780.1	4e-19	96	53/91
Pool 4	qt81a10.g1.b	LOC_Os01g63790.1	2e-33	140	90/135
Pool 4	Pool4.Contig97	LOC_Os01g63810.1	1e-54	97	58/129
Pool 4	Pool4.Contig46	LOC_Os01g63820.1	2e-11	68	35/43
Pool 4	Pool4.Contig95	LOC_Os01g63820.1	6e-22	74	34/36
Pool 4	Pool4.Contig101	LOC_Os01g63820.1	1e-78	292	144/158
Pool 4	ix32h09.b1.b	LOC_Os01g63830.1	9e-69	257	134/154
Pool 4	Pool4.Contig91	LOC_Os01g63840.1	4e-16	84	43/62
Pool 4	Pool4.Contig108	LOC_Os01g63840.1	2e-15	83	43/66
Pool 4	Pool4.Contig67	LOC_Os01g63854.1	5e-52	119	57/65
Pool 4	Pool4.Contig109	LOC_Os01g63854.1	6e-45	180	89/106
Pool 4	ix33d07.g1.b	LOC_Os01g63870.1	3e-64	196	90/97
Pool 4	Pool4.Contig26	LOC_Os01g63870.1	2e-31	134	64/73
Pool 4	Pool4.Contig120	LOC_Os01g63910.1	5e-38	158	100/256
Pool 4	Pool4.Contig87	LOC_Os01g63910.2	1e-64	245	200/586
Pool 4	Pool4.Contig29	LOC_Os01g63940.1	1e-24	111	71/163

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 4	Pool4.Contig119	LOC_Os01g63940.1	4e-22	105	64/149
Pool 4	Pool4.Contig105	LOC_Os01g63950.1	1e-44	126	68/100
Pool 4	Pool4.Contig143	LOC_Os01g63950.1	1e-72	273	135/173
Pool 4	Pool4.Contig112	LOC_Os01g63970.1	4e-47	187	85/104
Pool 4	Pool4.Contig146	LOC_Os01g63980.1	8e-76	285	130/144
Pool 4	Pool4.Contig44	LOC_Os01g64000.1	5e-42	140	75/100
Pool 4	Pool4.Contig145	LOC_Os01g64010.2	6e-92	338	160/183
Pool 4	qt80a08.g1.b	LOC_Os01g64020.2	1e-50	196	106/164
Pool 4	Pool4.Contig61	LOC_Os01g64050.1	1e-38	158	132/407
Pool 4	Pool4.Contig22	LOC_Os01g65780.3	3e-20	97	50/109
Pool 4	Pool4.Contig130	LOC_Os02g17260.1	1e-11	70	45/133
Pool 4	Pool4.Contig75	LOC_Os02g20430.1	1e-49	151	72/79
Pool 4	Pool4.Contig136	LOC_Os03g06190.1	6e-24	111	73/193
Pool 4	Pool4.Contig147	LOC_Os03g08500.1	8e-35	149	106/223
Pool 4	qt81b05.g1.b	LOC_Os03g46740.1	5e-20	95	56/168
Pool 4	Pool4.Contig142	LOC_Os03g55260.1	e-158	293	155/315
Pool 4	Pool4.Contig19	LOC_Os04g55290.2	1e-24	108	49/50
Pool 4	Pool4.Contig47	LOC_Os05g11770.1	2e-15	82	69/208
Pool 4	Pool4.Contig111	LOC_Os05g11770.1	1e-25	115	57/84
Pool 4	Pool4.Contig126	LOC_Os05g11770.1	7e-20	97	56/111
Pool 4	Pool4.Contig79	LOC_Os05g37340.1	2e-35	148	105/231
Pool 4	Pool4.Contig110	LOC_Os05g37390.2	4e-26	118	49/52
Pool 4	Pool4.Contig18	LOC_Os06g22700.1	1e-74	277	140/221
Pool 4	Pool4.Contig107	LOC_Os07g27140.1	2e-52	178	114/281
Pool 4	Pool4.Contig121	LOC_Os07g27140.1	4e-16	86	42/61
Pool 4	Pool4.Contig36	LOC_Os07g30810.1	2e-15	81	45/87
Pool 4	Pool4.Contig08	LOC_Os12g22400.1	2e-22	104	76/272
Pool 5	ix41b01.g1.b	LOC_Os01g16900.1	1e-21	88	46/104
Pool 5	Pool5.Contig54	LOC_Os01g26160.1	5e-18	91	57/168
Pool 5	Pool5.Contig36	LOC_Os01g51220.1	1e-17	88	50/81
Pool 5	Pool5.Contig03	LOC_Os01g63990.4	2e-26	118	56/75
Pool 5	Pool5.Contig16	LOC_Os01g64000.1	4e-37	151	80/105
Pool 5	Pool5.Contig129	LOC_Os01g64000.1	9e-56	217	123/194
Pool 5	Pool5.Contig05	LOC_Os01g64010.2	3e-61	232	104/116
Pool 5	Pool5.Contig48	LOC_Os01g64010.2	2e-30	130	64/76
Pool 5	Pool5.Contig128	LOC_Os01g64010.2	2e-29	129	65/76
Pool 5	Pool5.Contig11	LOC_Os01g64020.1	2e-20	98	47/52
Pool 5	Pool5.Contig56	LOC_Os01g64020.2	2e-73	273	144/211
Pool 5	Pool5.Contig125	LOC_Os01g64030.4	0.0	409	207/227
Pool 5	Pool5.Contig142	LOC_Os01g64100.1	e-133	263	127/166
Pool 5	Pool5.Contig148	LOC_Os01g64100.1	e-140	282	133/164
Pool 5	Pool5.Contig149	LOC_Os01g64100.1	3e-63	241	109/134
Pool 5	Pool5.Contig154	LOC_Os01g64120.1	4e-52	206	113/171
Pool 5	Pool5.Contig155	LOC_Os01g64180.3	1e-75	284	159/277
Pool 5	Pool5.Contig33	LOC_Os01g64250.1	1e-68	256	133/153
Pool 5	Pool5.Contig72	LOC_Os01g64250.1	3e-62	236	119/144
Pool 5	Pool5.Contig99	LOC_Os01g64250.1	4e-19	94	47/52
Pool 5	Pool5.Contig131	LOC_Os01g64270.1	1e-17	90	49/81
Pool 5	Pool5.Contig152	LOC_Os01g64300.1	2e-84	313	165/219
Pool 5	Pool5.Contig143	LOC_Os01g64310.1	3e-17	89	57/117
Pool 5	Pool5.Contig145	LOC_Os01g64310.1	2e-35	149	85/146

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 5	Pool5.Contig97	LOC_Os01g64380.1	2e-20	99	55/77
Pool 5	ix38h10.b1.b	LOC_Os01g64410.1	8e-70	261	160/242
Pool 5	Pool5.Contig66	LOC_Os01g64430.1	1e-32	139	86/119
Pool 5	Pool5.Contig141	LOC_Os01g64440.1	3e-33	141	92/170
Pool 5	Pool5.Contig150	LOC_Os01g64450.1	2e-86	319	170/243
Pool 5	Pool5.Contig153	LOC_Os01g64490.1	0.0	408	227/412
Pool 5	ix42h02.g1.b	LOC_Os01g64520.1	3e-50	166	89/137
Pool 5	Pool5.Contig22	LOC_Os01g64540.1	5e-55	212	105/156
Pool 5	Pool5.Contig40	LOC_Os01g64540.1	e-115	413	201/278
Pool 5	Pool5.Contig52	LOC_Os01g64540.1	4e-18	91	41/47
Pool 5	Pool5.Contig78	LOC_Os01g64570.1	6e-33	139	73/98
Pool 5	Pool5.Contig96	LOC_Os01g64570.1	e-109	253	134/176
Pool 5	ix40c04.g1.b	LOC_Os01g64590.1	2e-42	169	84/111
Pool 5	ix38g02.b1.b	LOC_Os01g64630.1	5e-87	289	142/151
Pool 5	Pool5.Contig114	LOC_Os01g64650.2	5e-30	130	68/78
Pool 5	ix41e05.b1.b	LOC_Os01g64650.3	7e-11	65	31/39
Pool 5	ix38a08.g1.b	LOC_Os01g64660.1	7e-46	127	63/69
Pool 5	Pool5.Contig111	LOC_Os01g64670.2	1e-59	115	61/100
Pool 5	ix42d05.g1.b	LOC_Os01g64680.1	5e-29	125	64/75
Pool 5	ix40c08.b1.b	LOC_Os01g64690.1	4e-16	83	43/44
Pool 5	Pool5.Contig132	LOC_Os02g21660.1	8e-34	144	70/74
Pool 5	Pool5.Contig15	LOC_Os02g24230.1	2e-19	94	49/146
Pool 5	Pool5.Contig156	LOC_Os02g33000.1	e-112	237	115/212
Pool 5	ix40f04.b1.b	LOC_Os02g41470.3	1e-35	146	78/185
Pool 5	Pool5.Contig26	LOC_Os02g44132.1	1e-53	208	105/126
Pool 5	ix41b01.b1.b	LOC_Os03g07850.1	4e-16	82	41/96
Pool 5	Pool5.Contig134	LOC_Os04g13750.1	2e-94	345	197/507
Pool 5	Pool5.Contig107	LOC_Os05g18170.1	1e-11	70	32/68
Pool 5	Pool5.Contig04	LOC_Os05g36270.1	3e-28	80	37/43
Pool 5	Pool5.Contig118	LOC_Os05g36310.1	1e-26	120	68/143
Pool 5	ix42b05.b1.b	LOC_Os06g41560.1	1e-17	86	48/105
Pool 5	Pool5.Contig135	LOC_Os06g41560.1	e-157	555	308/561
Pool 5	Pool5.Contig100	LOC_Os08g44380.1	1e-35	134	67/71
Pool 5	ix38f06.b1.b	LOC_Os11g31430.1	6e-23	68	35/71
Pool 5	Pool5.Contig58	LOC_Os11g31470.1	3e-12	64	40/99
Pool 5	qt82d11.b1.b	LOC_Os12g39860.1	8e-47	184	87/157
Pool 6	qt83f04.b1.b	LOC_Os01g03020.2	9e-51	197	109/255
Pool 6	Pool6.Contig97	LOC_Os01g27990.1	9e-30	129	61/82
Pool 6	Pool6.Contig131	LOC_Os01g27990.2	e-164	578	293/337
Pool 6	ix49b02.g1.b	LOC_Os01g56900.2	4e-66	248	129/230
Pool 6	Pool6.Contig128	LOC_Os01g64630.1	5e-22	80	37/38
Pool 6	Pool6.Contig48	LOC_Os01g64650.2	2e-25	113	58/68
Pool 6	ix49h09.b1.b	LOC_Os01g64660.2	9e-31	129	64/70
Pool 6	Pool6.Contig43	LOC_Os01g64660.2	1e-27	65	29/32
Pool 6	Pool6.Contig75	LOC_Os01g64660.2	1e-20	65	30/36
Pool 6	Pool6.Contig95	LOC_Os01g64660.2	2e-15	79	35/43
Pool 6	Pool6.Contig154	LOC_Os01g64670.2	1e-39	97	69/157
Pool 6	ix49a02.g1.b	LOC_Os01g64680.1	2e-29	127	65/75
Pool 6	Pool6.Contig122	LOC_Os01g64690.1	7e-15	81	42/43
Pool 6	Pool6.Contig84	LOC_Os01g64700.1	5e-60	229	108/122
Pool 6	Pool6.Contig160	LOC_Os01g64700.1	e-124	446	224/283

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 6	Pool6.Contig118	LOC_Os01g64720.2	2e-18	91	42/54
Pool 6	Pool6.Contig69	LOC_Os01g64730.2	4e-64	218	124/183
Pool 6	Pool6.Contig62	LOC_Os01g64750.1	4e-25	114	53/66
Pool 6	Pool6.Contig148	LOC_Os01g64750.1	3e-31	135	69/123
Pool 6	Pool6.Contig169	LOC_Os01g64750.1	4e-25	115	54/66
Pool 6	Pool6.Contig175	LOC_Os01g64750.1	8e-29	129	66/82
Pool 6	Pool6.Contig167	LOC_Os01g64760.1	2e-78	293	145/163
Pool 6	Pool6.Contig174	LOC_Os01g64780.1	6e-26	116	55/58
Pool 6	Pool6.Contig94	LOC_Os01g64790.1	2e-48	191	98/131
Pool 6	Pool6.Contig134	LOC_Os01g64820.1	7e-63	239	126/175
Pool 6	Pool6.Contig162	LOC_Os01g64820.1	2e-27	123	69/107
Pool 6	Pool6.Contig171	LOC_Os01g64820.1	e-115	288	163/321
Pool 6	Pool6.Contig145	LOC_Os01g64850.1	2e-35	149	80/132
Pool 6	Pool6.Contig59	LOC_Os01g64870.1	e-111	400	197/231
Pool 6	Pool6.Contig121	LOC_Os01g64870.1	3e-34	114	59/83
Pool 6	Pool6.Contig172	LOC_Os01g64890.1	5e-41	169	81/100
Pool 6	Pool6.Contig89	LOC_Os01g64900.1	3e-31	134	80/144
Pool 6	Pool6.Contig46	LOC_Os01g64930.1	1e-22	105	50/53
Pool 6	Pool6.Contig80	LOC_Os01g64930.1	2e-44	177	83/101
Pool 6	Pool6.Contig123	LOC_Os01g64960.1	2e-73	150	72/88
Pool 6	Pool6.Contig140	LOC_Os01g64970.1	3e-25	114	62/98
Pool 6	Pool6.Contig144	LOC_Os01g64970.1	3e-75	187	93/118
Pool 6	Pool6.Contig158	LOC_Os01g64980.1	e-161	421	229/316
Pool 6	Pool6.Contig68	LOC_Os01g64990.2	1e-36	151	86/143
Pool 6	Pool6.Contig65	LOC_Os01g65000.1	8e-15	79	36/47
Pool 6	Pool6.Contig96	LOC_Os01g65000.1	e-105	380	176/188
Pool 6	Pool6.Contig107	LOC_Os01g65070.1	1e-19	91	39/41
Pool 6	Pool6.Contig152	LOC_Os01g65070.1	5e-29	128	63/78
Pool 6	Pool6.Contig90	LOC_Os01g65080.2	1e-97	354	185/263
Pool 6	Pool6.Contig165	LOC_Os01g65080.2	e-135	480	225/255
Pool 6	Pool6.Contig83	LOC_Os01g65110.1	e-113	408	207/241
Pool 6	Pool6.Contig33	LOC_Os01g65130.1	5e-20	97	48/61
Pool 6	Pool6.Contig119	LOC_Os01g65150.1	1e-85	308	147/226
Pool 6	Pool6.Contig12	LOC_Os01g65169.1	9e-36	147	69/95
Pool 6	Pool6.Contig39	LOC_Os01g65169.1	2e-33	140	79/135
Pool 6	Pool6.Contig61	LOC_Os01g65169.1	3e-27	109	55/73
Pool 6	Pool6.Contig114	LOC_Os01g65169.1	3e-41	167	98/181
Pool 6	Pool6.Contig26	LOC_Os01g65210.1	e-102	210	109/182
Pool 6	Pool6.Contig28	LOC_Os01g65210.1	6e-22	103	45/70
Pool 6	Pool6.Contig132	LOC_Os01g65230.1	7e-38	157	90/154
Pool 6	ix49h07.b1.b	LOC_Os01g65260.1	1e-77	286	143/152
Pool 6	Pool6.Contig103	LOC_Os01g65260.1	1e-81	302	163/223
Pool 6	qt83b12.g1.b	LOC_Os01g65310.1	6e-27	118	60/99
Pool 6	Pool6.Contig45	LOC_Os01g65310.1	1e-16	86	43/46
Pool 6	Pool6.Contig17	LOC_Os01g65320.1	3e-30	130	66/101
Pool 6	Pool6.Contig142	LOC_Os01g65320.1	0.0	670	360/524
Pool 6	Pool6.Contig72	LOC_Os01g65330.1	4e-13	74	35/42
Pool 6	Pool6.Contig115	LOC_Os01g65330.1	2e-99	301	148/168
Pool 6	Pool6.Contig137	LOC_Os01g65330.2	2e-46	186	100/109
Pool 6	Pool6.Contig15	LOC_Os02g49170.1	2e-15	81	36/59
Pool 6	ix46f04.g1.b	LOC_Os02g58340.1	2e-17	87	62/225

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 6	Pool6.Contig113	LOC_Os03g24640.1	3e-27	120	55/74
Pool 6	Pool6.Contig44	LOC_Os03g24760.1	1e-23	99	43/60
Pool 6	Pool6.Contig149	LOC_Os03g50890.1	e-157	233	113/123
Pool 6	ix49f08.b1.b	LOC_Os04g48770.1	2e-32	108	61/119
Pool 6	ix45d06.g1.b	LOC_Os05g21200.1	4e-44	175	87/197
Pool 6	Pool6.Contig88	LOC_Os05g35060.1	6e-37	78	38/63
Pool 6	Pool6.Contig166	LOC_Os05g35730.1	3e-24	76	30/42
Pool 6	Pool6.Contig156	LOC_Os06g17840.1	2e-24	56	40/127
Pool 6	Pool6.Contig136	LOC_Os06g22840.1	3e-12	49	35/86
Pool 6	Pool6.Contig117	LOC_Os06g42770.1	e-151	345	174/211
Pool 6	Pool6.Contig31	LOC_Os07g15440.1	3e-38	156	89/211
Pool 6	qt83b01.b1.b	LOC_Os07g46460.1	5e-41	165	86/170
Pool 6	Pool6.Contig170	LOC_Os08g23730.1	7e-11	69	36/55
Pool 6	Pool6.Contig146	LOC_Os11g24110.1	7e-63	133	62/130
Pool 6	Pool6.Contig66	LOC_Os12g32200.1	3e-26	118	100/278
Pool 7	Pool7.Contig213	LOC_Os01g33490.1	8e-93	340	182/276
Pool 7	ix54f10.b1.b	LOC_Os01g65150.1	4e-71	262	132/205
Pool 7	ix56e08.g1.b	LOC_Os01g65210.1	4e-53	160	82/142
Pool 7	Pool7.Contig64	LOC_Os01g65230.1	8e-24	108	51/55
Pool 7	Pool7.Contig134	LOC_Os01g65230.1	2e-44	137	66/74
Pool 7	ix54b12.g1.b	LOC_Os01g65260.1	2e-41	167	91/143
Pool 7	ix54a11.g1.b	LOC_Os01g65310.1	5e-17	86	43/46
Pool 7	ix57c09.g1.b	LOC_Os01g65320.1	4e-25	112	50/60
Pool 7	qt84e06.b1.b	LOC_Os01g65320.1	3e-14	77	36/38
Pool 7	Pool7.Contig82	LOC_Os01g65320.1	1e-45	181	82/85
Pool 7	ix55a08.b1.b	LOC_Os01g65330.1	3e-11	65	30/34
Pool 7	ix55a08.g1.b	LOC_Os01g65330.1	4e-68	183	89/100
Pool 7	Pool7.Contig45	LOC_Os01g65350.1	2e-63	240	116/125
Pool 7	Pool7.Contig87	LOC_Os01g65350.1	1e-24	111	55/71
Pool 7	ix55f11.b1.b	LOC_Os01g65350.2	1e-67	252	120/127
Pool 7	ix55g05.b1.b	LOC_Os01g65370.1	7e-20	94	50/73
Pool 7	ix53b10.g1.b	LOC_Os01g65380.1	1e-39	160	85/121
Pool 7	ix53c06.b1.b	LOC_Os01g65380.1	2e-40	163	79/83
Pool 7	ix54b05.b1.b	LOC_Os01g65380.1	7e-90	327	163/180
Pool 7	Pool7.Contig120	LOC_Os01g65400.4	4e-35	146	73/86
Pool 7	ix54f05.g1.b	LOC_Os01g65410.1	4e-56	215	105/120
Pool 7	ix58h08.b1.b	LOC_Os01g65410.1	1e-16	82	38/47
Pool 7	Pool7.Contig121	LOC_Os01g65410.1	3e-28	124	57/58
Pool 7	ix55d04.b1.b	LOC_Os01g65420.1	2e-29	124	75/109
Pool 7	Pool7.Contig184	LOC_Os01g65440.1	1e-35	149	77/113
Pool 7	Pool7.Contig171	LOC_Os01g65450.1	3e-23	107	49/67
Pool 7	Pool7.Contig178	LOC_Os01g65450.1	9e-19	85	38/52
Pool 7	Pool7.Contig210	LOC_Os01g65450.1	4e-34	145	73/99
Pool 7	Pool7.Contig229	LOC_Os01g65460.1	1e-32	112	57/117
Pool 7	Pool7.Contig231	LOC_Os01g65460.1	e-102	199	91/118
Pool 7	Pool7.Contig159	LOC_Os01g65480.1	4e-70	263	126/146
Pool 7	Pool7.Contig132	LOC_Os01g65490.1	3e-90	330	155/214
Pool 7	ix53b09.g1.b	LOC_Os01g65500.1	1e-16	85	50/83
Pool 7	Pool7.Contig162	LOC_Os01g65500.1	e-102	347	185/282
Pool 7	ix57g06.b1.b	LOC_Os01g65510.1	4e-38	154	71/103
Pool 7	Pool7.Contig183	LOC_Os01g65520.1	e-131	466	252/397

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 7	Pool7.Contig223	LOC_Os01g65520.1	3e-13	75	44/71
Pool 7	Pool7.Contig57	LOC_Os01g65550.2	5e-20	96	44/45
Pool 7	Pool7.Contig42	LOC_Os01g65560.1	9e-19	92	42/48
Pool 7	Pool7.Contig43	LOC_Os01g65560.1	5e-35	145	79/108
Pool 7	Pool7.Contig104	LOC_Os01g65560.1	8e-12	69	42/89
Pool 7	ix56e06.g1.b	LOC_Os01g65580.1	3e-21	87	43/44
Pool 7	Pool7.Contig41	LOC_Os01g65580.1	2e-22	103	53/55
Pool 7	Pool7.Contig187	LOC_Os01g65580.1	3e-24	111	62/87
Pool 7	Pool7.Contig226	LOC_Os01g65580.1	1e-24	113	68/121
Pool 7	Pool7.Contig170	LOC_Os01g65590.1	7e-14	76	43/75
Pool 7	Pool7.Contig186	LOC_Os01g65590.1	5e-74	153	77/114
Pool 7	Pool7.Contig190	LOC_Os01g65600.1	0.0	465	221/272
Pool 7	Pool7.Contig234	LOC_Os01g65630.2	2e-30	133	78/123
Pool 7	Pool7.Contig189	LOC_Os01g65650.1	e-137	486	243/301
Pool 7	Pool7.Contig217	LOC_Os01g65650.1	e-153	540	263/318
Pool 7	qt84h09.g1.b	LOC_Os01g65660.1	2e-54	208	107/158
Pool 7	Pool7.Contig199	LOC_Os01g65670.1	9e-94	182	95/121
Pool 7	Pool7.Contig238	LOC_Os01g65670.1	2e-37	157	80/87
Pool 7	qt84c09.g1.b	LOC_Os01g65680.1	6e-56	214	96/129
Pool 7	Pool7.Contig73	LOC_Os01g65680.1	7e-33	138	78/155
Pool 7	Pool7.Contig188	LOC_Os01g65690.1	2e-48	119	53/84
Pool 7	ix56b04.b1.b	LOC_Os01g65700.1	1e-67	254	130/192
Pool 7	Pool7.Contig48	LOC_Os01g65720.1	5e-32	136	66/91
Pool 7	Pool7.Contig50	LOC_Os01g65720.1	e-134	478	224/256
Pool 7	Pool7.Contig107	LOC_Os01g65730.1	5e-67	251	125/171
Pool 7	Pool7.Contig29	LOC_Os01g65740.1	2e-38	158	93/159
Pool 7	ix53d10.g1.b	LOC_Os01g65780.2	4e-29	125	67/156
Pool 7	Pool7.Contig192	LOC_Os01g65780.2	e-105	382	185/256
Pool 7	Pool7.Contig105	LOC_Os01g65780.3	0.0	677	316/345
Pool 7	ix53a01.g1.b	LOC_Os01g65790.1	8e-87	317	160/189
Pool 7	Pool7.Contig100	LOC_Os01g65800.1	2e-44	177	82/94
Pool 7	Pool7.Contig164	LOC_Os01g65830.1	e-150	358	175/189
Pool 7	Pool7.Contig46	LOC_Os01g65840.1	e-141	498	244/330
Pool 7	Pool7.Contig143	LOC_Os01g65840.1	3e-55	214	103/130
Pool 7	Pool7.Contig167	LOC_Os01g65850.1	2e-45	103	47/60
Pool 7	Pool7.Contig220	LOC_Os01g65850.1	2e-34	106	49/64
Pool 7	Pool7.Contig232	LOC_Os01g65850.1	8e-49	194	96/150
Pool 7	Pool7.Contig237	LOC_Os01g65850.1	4e-89	328	169/203
Pool 7	Pool7.Contig67	LOC_Os01g65880.1	5e-16	84	42/50
Pool 7	Pool7.Contig122	LOC_Os01g65880.1	1e-17	88	57/88
Pool 7	Pool7.Contig235	LOC_Os01g65890.1	6e-93	341	190/335
Pool 7	Pool7.Contig37	LOC_Os01g65920.2	7e-37	152	73/89
Pool 7	Pool7.Contig55	LOC_Os01g65920.2	2e-29	127	60/87
Pool 7	Pool7.Contig101	LOC_Os01g65920.2	1e-93	301	156/220
Pool 7	ix56e04.g1.b	LOC_Os01g65940.1	6e-84	248	120/132
Pool 7	Pool7.Contig208	LOC_Os01g65940.1	1e-26	89	43/69
Pool 7	Pool7.Contig34	LOC_Os01g65960.1	2e-16	73	38/57
Pool 7	ix53d04.g1.b	LOC_Os01g65986.1	8e-22	73	45/75
Pool 7	ix54e06.b1.b	LOC_Os01g66000.1	8e-15	78	48/107
Pool 7	Pool7.Contig90	LOC_Os01g66000.1	8e-24	108	60/99
Pool 7	ix56d08.b1.b	LOC_Os01g66010.1	3e-36	149	69/75

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 7	Pool7.Contig109	LOC_Os01g66020.1	e-141	498	259/349
Pool 7	Pool7.Contig216	LOC_Os01g66030.1	2e-18	93	42/45
Pool 7	Pool7.Contig200	LOC_Os01g66050.1	3e-27	121	66/94
Pool 7	Pool7.Contig205	LOC_Os01g66050.1	2e-23	105	60/100
Pool 7	Pool7.Contig202	LOC_Os01g68550.1	6e-32	137	71/93
Pool 7	Pool7.Contig228	LOC_Os01g68550.1	2e-58	226	108/150
Pool 7	qt84e03.b1.b	LOC_Os02g11940.1	2e-15	70	35/57
Pool 7	ix56e05.g1.b	LOC_Os02g42200.2	4e-35	145	75/112
Pool 7	ix56b08.b1.b	LOC_Os04g48820.1	1e-21	101	66/212
Pool 7	ix55c08.g1.b	LOC_Os05g34900.1	2e-16	63	30/42
Pool 7	ix53f11.g1.b	LOC_Os05g34980.1	1e-56	216	117/161
Pool 7	Pool7.Contig85	LOC_Os05g35060.1	4e-32	79	35/38
Pool 7	Pool7.Contig116	LOC_Os05g35060.1	1e-52	79	38/66
Pool 7	Pool7.Contig96	LOC_Os05g35140.1	1e-16	84	40/58
Pool 7	Pool7.Contig31	LOC_Os05g35160.1	6e-52	202	125/238
Pool 7	ix57e09.g1.b	LOC_Os05g35274.3	1e-37	152	72/88
Pool 7	Pool7.Contig79	LOC_Os06g42770.1	2e-50	196	113/237
Pool 7	Pool7.Contig212	LOC_Os06g42770.1	4e-86	280	134/179
Pool 7	ix53f12.b1.b	LOC_Os08g07810.1	4e-71	264	129/181
Pool 7	ix53f11.b1.b	LOC_Os08g09770.1	2e-78	288	148/182
Pool 7	ix55c03.g1.b	LOC_Os08g21330.2	1e-13	74	49/165
Pool 7	Pool7.Contig198	LOC_Os09g10650.1	1e-72	272	140/197
Pool 7	Pool7.Contig106	LOC_Os10g11340.1	e-115	413	196/252
Pool 7	Pool7.Contig206	LOC_Os10g18470.1	2e-30	132	77/118
Pool 7	ix55c11.g1.b	LOC_Os11g24110.1	7e-63	238	107/225
Pool 7	ix54a03.g1.b	LOC_Os11g34230.1	3e-15	61	42/103
Pool 7	Pool7.Contig38	LOC_Os12g13440.1	e-107	385	228/451
Pool 7	Pool7.Contig40	LOC_Os12g13440.1	5e-43	172	85/107
Pool 7	Pool7.Contig224	LOC_Os12g13440.1	0.0	528	304/576
Pool 7	ix53d12.b1.b	LOC_Os12g32200.1	5e-21	99	48/72
Pool 7	ix54h03.g1.b	LOC_Os12g32200.1	2e-24	110	86/232
Pool 7	ix55e11.g1.b	LOC_Os12g32986.1	3e-40	162	82/153
Pool 7	ix55h02.g1.b	LOC_Os12g37960.1	2e-19	94	55/141
Pool 8	ix66c05.b1.b	LOC_Os01g03020.1	3e-23	98	51/128
Pool 8	ix62h04.g1.b	LOC_Os01g18070.1	6e-16	82	50/172
Pool 8	ix62b02.g1.b	LOC_Os01g24680.1	3e-21	100	59/172
Pool 8	Pool8.Contig172	LOC_Os01g42400.1	2e-36	152	75/111
Pool 8	Pool8.Contig127	LOC_Os01g45880.2	8e-22	92	42/69
Pool 8	Pool8.Contig148	LOC_Os01g65010.1	e-178	412	208/343
Pool 8	Pool8.Contig176	LOC_Os01g66030.2	4e-28	124	62/63
Pool 8	Pool8.Contig165	LOC_Os01g66050.1	9e-23	105	60/100
Pool 8	Pool8.Contig188	LOC_Os01g66050.1	1e-35	150	91/148
Pool 8	Pool8.Contig162	LOC_Os01g66070.1	4e-39	161	82/165
Pool 8	Pool8.Contig186	LOC_Os01g66100.1	e-129	460	239/341
Pool 8	Pool8.Contig184	LOC_Os01g66110.2	3e-74	278	133/185
Pool 8	Pool8.Contig93	LOC_Os01g66120.3	3e-49	192	89/92
Pool 8	Pool8.Contig182	LOC_Os01g66120.3	5e-52	204	114/165
Pool 8	Pool8.Contig192	LOC_Os01g66130.1	e-175	525	292/492
Pool 8	Pool8.Contig190	LOC_Os01g66140.1	2e-19	96	59/148
Pool 8	Pool8.Contig196	LOC_Os01g66140.1	1e-80	265	145/285
Pool 8	Pool8.Contig160	LOC_Os01g66150.1	5e-32	137	80/127

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 8	Pool8.Contig169	LOC_Os01g66150.1	2e-13	75	33/43
Pool 8	ix62h07.b1.b	LOC_Os01g66160.1	6e-71	264	137/172
Pool 8	ix65f07.b1.b	LOC_Os01g66170.1	2e-11	67	33/36
Pool 8	Pool8.Contig84	LOC_Os01g66170.1	3e-20	97	46/51
Pool 8	ix66b04.g1.b	LOC_Os01g66190.2	2e-38	156	76/94
Pool 8	ix64f02.g1.b	LOC_Os01g66200.1	5e-54	208	116/172
Pool 8	qt85f08.b1.b	LOC_Os01g66200.1	6e-22	102	47/53
Pool 8	ix62d05.g1.b	LOC_Os01g66260.1	4e-50	195	120/259
Pool 8	ix65f06.g1.b	LOC_Os01g66260.1	8e-51	181	88/155
Pool 8	ix66a02.b1.b	LOC_Os01g66280.1	3e-12	70	34/59
Pool 8	Pool8.Contig163	LOC_Os01g66280.1	e-108	391	187/223
Pool 8	Pool8.Contig154	LOC_Os01g66290.2	5e-14	78	55/118
Pool 8	ix62d11.g1.b	LOC_Os01g66300.1	9e-31	82	40/42
Pool 8	ix64g04.g1.b	LOC_Os01g66330.1	1e-44	152	80/119
Pool 8	Pool8.Contig69	LOC_Os01g66330.1	2e-65	246	126/179
Pool 8	Pool8.Contig70	LOC_Os01g66330.1	2e-17	87	54/87
Pool 8	ix65e02.g1.b	LOC_Os01g66350.1	5e-42	81	40/41
Pool 8	Pool8.Contig150	LOC_Os01g66350.1	4e-19	68	39/56
Pool 8	Pool8.Contig151	LOC_Os01g66350.1	2e-31	105	55/73
Pool 8	ix63h05.g1.b	LOC_Os01g66360.1	6e-15	79	59/155
Pool 8	Pool8.Contig136	LOC_Os01g66379.1	1e-28	126	87/158
Pool 8	ix65e03.g1.b	LOC_Os01g66490.1	2e-31	134	89/167
Pool 8	ix65c08.b1.b	LOC_Os01g66510.1	3e-11	67	30/35
Pool 8	Pool8.Contig120	LOC_Os01g66520.1	2e-43	136	77/139
Pool 8	Pool8.Contig131	LOC_Os01g66520.1	2e-35	149	82/153
Pool 8	ix64a06.g1.b	LOC_Os01g66560.1	5e-34	142	76/111
Pool 8	Pool8.Contig122	LOC_Os01g66560.1	2e-31	135	70/94
Pool 8	Pool8.Contig180	LOC_Os01g66560.1	e-142	464	235/297
Pool 8	ix65f05.g1.b	LOC_Os01g66580.1	2e-41	166	88/145
Pool 8	Pool8.Contig132	LOC_Os01g66590.3	7e-69	259	126/137
Pool 8	qt85b07.b1.b	LOC_Os01g66600.1	5e-77	285	161/232
Pool 8	Pool8.Contig185	LOC_Os01g66600.1	2e-31	87	46/85
Pool 8	ix65f11.g1.b	LOC_Os01g66690.1	1e-56	216	111/164
Pool 8	ix66a08.b1.b	LOC_Os01g66690.1	1e-15	80	41/66
Pool 8	qt85d09.b1.b	LOC_Os01g66690.1	5e-36	147	78/137
Pool 8	Pool8.Contig30	LOC_Os01g66690.1	3e-56	216	121/192
Pool 8	Pool8.Contig125	LOC_Os01g66690.1	8e-79	286	148/211
Pool 8	Pool8.Contig25	LOC_Os01g66700.1	5e-39	159	74/85
Pool 8	Pool8.Contig90	LOC_Os01g66720.1	7e-55	213	100/126
Pool 8	Pool8.Contig135	LOC_Os01g66720.1	3e-30	130	68/86
Pool 8	Pool8.Contig155	LOC_Os01g66830.1	3e-35	80	40/91
Pool 8	Pool8.Contig77	LOC_Os01g66840.1	7e-17	64	27/40
Pool 8	ix63a07.g1.b	LOC_Os01g66860.2	2e-25	113	52/60
Pool 8	ix66h07.g1.b	LOC_Os01g66890.2	3e-41	166	75/93
Pool 8	ix63b01.g1.b	LOC_Os01g66890.3	3e-26	114	54/85
Pool 8	Pool8.Contig35	LOC_Os01g66920.1	4e-78	288	136/162
Pool 8	Pool8.Contig38	LOC_Os01g66920.1	2e-30	131	61/78
Pool 8	qt85c04.b1.b	LOC_Os01g66960.1	6e-28	122	57/62
Pool 8	Pool8.Contig80	LOC_Os01g66960.1	4e-16	83	44/58
Pool 8	Pool8.Contig82	LOC_Os01g66970.1	5e-30	129	69/107
Pool 8	ix66b11.g1.b	LOC_Os01g66980.1	1e-25	74	34/60

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 8	ix66a09.b1.b	LOC_Os01g72340.1	5e-38	155	76/202
Pool 8	Pool8.Contig166	LOC_Os02g17260.1	2e-14	79	54/162
Pool 8	ix63h03.g1.b	LOC_Os02g49720.4	9e-25	110	52/104
Pool 8	Pool8.Contig31	LOC_Os02g54640.1	2e-15	80	56/145
Pool 8	ix62h04.b1.b	LOC_Os03g05780.1	4e-12	69	56/188
Pool 8	ix62d12.g1.b	LOC_Os03g49260.1	4e-91	323	145/184
Pool 8	qt85b07.g1.b	LOC_Os03g50510.1	3e-25	112	62/109
Pool 8	Pool8.Contig40	LOC_Os03g52860.1	e-102	219	101/124
Pool 8	ix62b05.g1.b	LOC_Os03g61920.1	9e-13	71	38/100
Pool 8	ix66e08.b1.b	LOC_Os04g28820.1	5e-44	172	83/103
Pool 8	Pool8.Contig36	LOC_Os04g28820.1	1e-77	268	123/194
Pool 8	ix64b01.b1.b	LOC_Os04g39520.1	1e-11	67	48/143
Pool 8	Pool8.Contig159	LOC_Os04g50700.1	2e-60	232	126/211
Pool 8	ix66b01.b1.b	LOC_Os05g34650.2	6e-19	92	62/156
Pool 8	ix62g04.g1.b	LOC_Os05g34770.1	1e-13	74	32/33
Pool 8	Pool8.Contig61	LOC_Os05g34770.1	7e-11	65	30/40
Pool 8	Pool8.Contig191	LOC_Os05g34900.1	2e-15	74	51/116
Pool 8	ix64c12.g1.b	LOC_Os06g07440.1	2e-24	109	68/165
Pool 8	ix63h12.b1.b	LOC_Os06g15990.2	6e-29	125	66/163
Pool 8	ix65b06.b1.b	LOC_Os07g10770.1	7e-11	65	49/133
Pool 8	Pool8.Contig71	LOC_Os07g17850.1	2e-74	276	132/250
Pool 8	ix64b05.g1.b	LOC_Os09g23810.1	1e-31	134	71/204
Pool 8	Pool8.Contig173	LOC_Os10g10730.1	0.0	632	318/463
Pool 8	ix66c07.g1.b	LOC_Os10g11620.1	3e-86	315	145/187
Pool 8	ix64g04.b1.b	LOC_Os11g09630.1	8e-29	88	37/55
Pool 8	Pool8.Contig119	LOC_Os12g10540.4	4e-26	117	55/64
Pool 9	Pool9.Contig149	LOC_Os01g42900.1	1e-35	147	84/150
Pool 9	Pool9.Contig205	LOC_Os01g42900.1	1e-70	266	133/243
Pool 9	ix69h06.b1.b	LOC_Os01g53810.1	6e-13	72	36/51
Pool 9	Pool9.Contig80	LOC_Os01g66940.1	7e-13	74	35/45
Pool 9	Pool9.Contig184	LOC_Os01g66940.1	2e-57	221	109/112
Pool 9	Pool9.Contig170	LOC_Os01g66970.1	2e-29	129	69/107
Pool 9	Pool9.Contig197	LOC_Os01g66970.2	1e-22	106	53/65
Pool 9	Pool9.Contig213	LOC_Os01g66980.1	1e-25	105	60/95
Pool 9	Pool9.Contig73	LOC_Os01g67000.1	8e-25	111	63/107
Pool 9	Pool9.Contig122	LOC_Os01g67000.1	1e-14	52	25/28
Pool 9	Pool9.Contig173	LOC_Os01g67000.1	4e-12	70	36/57
Pool 9	Pool9.Contig190	LOC_Os01g67010.1	3e-43	174	83/111
Pool 9	ix74a07.b1.b	LOC_Os01g67040.2	7e-39	157	87/119
Pool 9	ix71g02.g1.b	LOC_Os01g67090.1	1e-14	77	45/65
Pool 9	Pool9.Contig204	LOC_Os01g67090.1	4e-90	186	112/161
Pool 9	ix72f07.g1.b	LOC_Os01g67100.1	2e-12	70	39/87
Pool 9	Pool9.Contig19	LOC_Os01g67100.1	4e-65	245	135/196
Pool 9	Pool9.Contig179	LOC_Os01g67100.1	e-118	411	222/327
Pool 9	Pool9.Contig148	LOC_Os01g67110.1	8e-75	279	147/164
Pool 9	Pool9.Contig210	LOC_Os01g67120.1	2e-28	100	48/60
Pool 9	Pool9.Contig195	LOC_Os01g67120.2	2e-21	102	65/135
Pool 9	Pool9.Contig175	LOC_Os01g67134.4	8e-33	96	42/60
Pool 9	Pool9.Contig189	LOC_Os01g67134.6	7e-59	134	81/168
Pool 9	Pool9.Contig177	LOC_Os01g67170.1	1e-29	130	67/105
Pool 9	Pool9.Contig214	LOC_Os01g67170.1	e-166	566	271/346

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 9	Pool9.Contig51	LOC_Os01g67180.1	2e-19	93	44/54
Pool 9	Pool9.Contig70	LOC_Os01g67190.1	4e-24	107	49/78
Pool 9	qt86b04.b1.b	LOC_Os01g67210.1	7e-27	117	67/134
Pool 9	Pool9.Contig33	LOC_Os01g67210.1	3e-49	193	116/183
Pool 9	Pool9.Contig199	LOC_Os01g67220.3	2e-36	152	70/75
Pool 9	ix72b06.g1.b	LOC_Os01g67240.1	8e-40	160	80/89
Pool 9	Pool9.Contig215	LOC_Os01g67240.1	7e-87	321	161/176
Pool 9	ix70e12.b1.b	LOC_Os01g67250.1	4e-25	112	73/191
Pool 9	Pool9.Contig90	LOC_Os01g67250.1	1e-49	134	65/66
Pool 9	ix69g04.g1.b	LOC_Os01g67290.1	2e-30	130	71/127
Pool 9	ix70b12.b1.b	LOC_Os01g67290.1	6e-42	100	48/67
Pool 9	Pool9.Contig29	LOC_Os01g67290.1	4e-72	169	80/119
Pool 9	Pool9.Contig74	LOC_Os01g67310.1	6e-37	152	73/101
Pool 9	Pool9.Contig17	LOC_Os01g67330.2	2e-16	85	42/44
Pool 9	Pool9.Contig18	LOC_Os01g67330.2	3e-15	80	39/51
Pool 9	Pool9.Contig67	LOC_Os01g67340.2	1e-22	104	59/102
Pool 9	Pool9.Contig71	LOC_Os01g67340.2	4e-16	84	41/42
Pool 9	Pool9.Contig32	LOC_Os01g67420.1	5e-16	82	37/48
Pool 9	Pool9.Contig121	LOC_Os01g67420.1	5e-26	91	43/52
Pool 9	Pool9.Contig159	LOC_Os01g67420.1	e-103	373	205/337
Pool 9	Pool9.Contig178	LOC_Os01g67430.1	6e-83	306	171/269
Pool 9	Pool9.Contig206	LOC_Os01g67430.1	3e-24	112	66/133
Pool 9	qt86g10.g1.b	LOC_Os01g67480.1	1e-35	147	109/213
Pool 9	Pool9.Contig169	LOC_Os01g67480.1	8e-11	67	44/77
Pool 9	Pool9.Contig191	LOC_Os01g67490.2	5e-36	151	75/115
Pool 9	Pool9.Contig13	LOC_Os01g67500.2	8e-68	254	152/255
Pool 9	Pool9.Contig27	LOC_Os01g67500.2	e-110	397	206/254
Pool 9	ix72f05.b1.b	LOC_Os01g67500.3	e-109	391	183/196
Pool 9	ix69f07.b1.b	LOC_Os01g67510.1	4e-15	79	40/47
Pool 9	ix71f11.b1.b	LOC_Os01g67510.1	9e-49	191	95/130
Pool 9	Pool9.Contig134	LOC_Os01g67510.1	3e-40	164	93/154
Pool 9	ix74g12.b1.b	LOC_Os01g67520.1	1e-58	150	73/90
Pool 9	Pool9.Contig49	LOC_Os01g67520.1	3e-40	159	84/137
Pool 9	ix70a01.b1.b	LOC_Os01g67530.1	2e-13	73	37/42
Pool 9	Pool9.Contig09	LOC_Os01g67530.1	2e-14	77	40/56
Pool 9	Pool9.Contig102	LOC_Os01g67530.1	3e-87	308	162/229
Pool 9	Pool9.Contig157	LOC_Os01g67530.1	1e-43	108	50/50
Pool 9	Pool9.Contig110	LOC_Os01g67560.1	4e-24	110	54/62
Pool 9	ix73h11.g1.b	LOC_Os01g67570.1	2e-21	100	44/55
Pool 9	Pool9.Contig188	LOC_Os01g67570.1	2e-33	141	72/95
Pool 9	ix70g02.g1.b	LOC_Os01g67590.1	1e-19	95	46/55
Pool 9	ix72h11.b1.b	LOC_Os01g67590.1	2e-11	67	35/40
Pool 9	Pool9.Contig211	LOC_Os01g67590.1	9e-44	96	45/60
Pool 9	Pool9.Contig81	LOC_Os01g67600.1	9e-14	76	35/43
Pool 9	Pool9.Contig167	LOC_Os01g67600.1	5e-60	230	108/111
Pool 9	ix73c11.b1.b	LOC_Os01g67630.1	8e-26	114	56/68
Pool 9	Pool9.Contig137	LOC_Os01g67630.1	7e-67	252	128/146
Pool 9	Pool9.Contig28	LOC_Os01g67650.1	3e-96	350	169/179
Pool 9	ix71d01.b1.b	LOC_Os01g67720.1	2e-20	97	44/53
Pool 9	Pool9.Contig160	LOC_Os01g67720.3	4e-28	124	62/69
Pool 9	ix69b02.b1.b	LOC_Os01g67740.1	6e-39	114	65/112

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 9	ix69b02.g1.b	LOC_Os01g67740.1	2e-11	50	25/30
Pool 9	ix69g11.b1.b	LOC_Os01g67740.1	1e-27	120	68/112
Pool 9	ix71h08.b1.b	LOC_Os01g67740.1	1e-96	350	187/245
Pool 9	qt86g07.b1.b	LOC_Os01g67740.1	2e-52	202	108/144
Pool 9	Pool9.Contig43	LOC_Os01g67740.1	4e-18	89	42/42
Pool 9	qt86b11.b1.b	LOC_Os01g67750.1	2e-45	180	79/99
Pool 9	ix69d07.g1.b	LOC_Os01g67970.1	2e-15	52	23/27
Pool 9	ix74b05.b1.b	LOC_Os01g67970.1	5e-12	69	40/118
Pool 9	ix74c12.b1.b	LOC_Os01g67970.1	1e-68	256	115/150
Pool 9	ix74c12.g1.b	LOC_Os01g67970.1	2e-35	146	93/193
Pool 9	Pool9.Contig21	LOC_Os01g67970.1	1e-49	194	83/100
Pool 9	Pool9.Contig20	LOC_Os01g67980.1	4e-75	279	134/176
Pool 9	qt86h08.g1.b	LOC_Os02g05330.1	6e-28	100	59/92
Pool 9	qt86e07.g1.b	LOC_Os02g10430.1	3e-21	97	42/52
Pool 9	ix71e07.g1.b	LOC_Os02g26510.1	2e-78	290	135/227
Pool 9	Pool9.Contig208	LOC_Os03g18620.1	1e-40	166	88/114
Pool 9	Pool9.Contig96	LOC_Os03g33870.1	4e-38	155	72/140
Pool 9	Pool9.Contig129	LOC_Os03g49350.1	e-102	311	148/191
Pool 9	ix73g07.g1.b	LOC_Os03g52860.1	5e-21	98	44/58
Pool 9	Pool9.Contig201	LOC_Os03g56710.1	3e-46	173	133/326
Pool 9	Pool9.Contig207	LOC_Os04g13750.1	3e-95	348	218/597
Pool 9	ix72f11.b1.b	LOC_Os04g41200.1	1e-22	103	53/102
Pool 9	Pool9.Contig105	LOC_Os05g15370.1	4e-52	118	54/58
Pool 9	ix69b06.g1.b	LOC_Os05g26840.1	5e-16	82	58/173
Pool 9	ix69h05.g1.b	LOC_Os05g42130.1	6e-62	234	119/157
Pool 9	ix74f08.b1.b	LOC_Os06g12100.1	7e-40	100	49/86
Pool 9	Pool9.Contig138	LOC_Os06g27770.1	5e-54	138	63/81
Pool 9	Pool9.Contig88	LOC_Os06g41750.1	8e-13	73	35/51
Pool 9	Pool9.Contig114	LOC_Os06g41750.1	4e-16	83	44/61
Pool 9	Pool9.Contig212	LOC_Os06g41750.1	2e-51	203	107/163
Pool 9	ix71f04.g1.b	LOC_Os07g42000.1	2e-14	77	49/117
Pool 9	Pool9.Contig168	LOC_Os09g03210.1	2e-49	195	99/190
Pool 9	Pool9.Contig216	LOC_Os09g03210.1	1e-14	81	43/96
Pool 9	Pool9.Contig144	LOC_Os10g10730.1	0.0	943	457/656
Pool 9	Pool9.Contig116	LOC_Os10g34140.1	5e-15	60	31/86
Pool 9	ix71a01.b1.b	LOC_Os10g34540.1	5e-21	74	34/43
Pool 9	Pool9.Contig186	LOC_Os10g35020.1	7e-35	147	66/77
Pool 9	Pool9.Contig156	LOC_Os10g38090.1	9e-19	94	45/60
Pool 9	Pool9.Contig171	LOC_Os12g12880.1	2e-27	122	62/122
Pool 9	Pool9.Contig85	LOC_Os12g20410.1	1e-16	62	25/43
Pool 10	Pool10.Contig88	LOC_Os01g42260.2	3e-12	70	39/79
Pool 10	Pool10.Contig75	LOC_Os01g52760.2	7e-24	109	66/112
Pool 10	iy39g04.b1.b	LOC_Os01g63270.2	1e-21	101	52/106
Pool 10	iy40e07.b1.b	LOC_Os01g67630.1	9e-26	114	56/68
Pool 10	iy39d06.b1.b	LOC_Os01g67720.1	7e-15	68	31/41
Pool 10	iy42d10.g1.b	LOC_Os01g67720.1	1e-14	77	37/47
Pool 10	iy42d10.b1.b	LOC_Os01g67720.2	7e-14	74	35/38
Pool 10	qu05d05.b1.b	LOC_Os01g67720.2	3e-15	79	36/41
Pool 10	Pool10.Contig62	LOC_Os01g67720.4	1e-14	78	39/44
Pool 10	Pool10.Contig27	LOC_Os01g67740.1	2e-17	86	55/99
Pool 10	Pool10.Contig51	LOC_Os01g67740.1	2e-29	100	50/58

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 10	Pool10.Contig69	LOC_Os01g67740.1	5e-28	122	61/62
Pool 10	Pool10.Contig85	LOC_Os01g67740.1	5e-26	115	66/110
Pool 10	Pool10.Contig94	LOC_Os01g67740.1	3e-64	244	138/220
Pool 10	Pool10.Contig106	LOC_Os01g67740.1	e-109	391	220/322
Pool 10	iy43e10.b1.b	LOC_Os01g67750.1	6e-46	181	88/103
Pool 10	Pool10.Contig128	LOC_Os01g67770.1	1e-77	184	86/100
Pool 10	iy43f12.b1.b	LOC_Os01g67820.1	1e-79	293	153/234
Pool 10	Pool10.Contig144	LOC_Os01g67830.1	2e-14	57	29/62
Pool 10	iy43h10.b1.b	LOC_Os01g67870.1	1e-22	103	57/90
Pool 10	Pool10.Contig180	LOC_Os01g67870.1	3e-73	140	63/71
Pool 10	Pool10.Contig82	LOC_Os01g67880.1	1e-22	72	34/36
Pool 10	Pool10.Contig124	LOC_Os01g67880.1	1e-66	181	94/114
Pool 10	Pool10.Contig91	LOC_Os01g67920.1	8e-23	105	48/53
Pool 10	Pool10.Contig198	LOC_Os01g67920.1	3e-47	114	61/95
Pool 10	Pool10.Contig197	LOC_Os01g67950.1	1e-34	147	70/76
Pool 10	qu05b11.b1.b	LOC_Os01g67960.1	1e-54	211	102/141
Pool 10	Pool10.Contig21	LOC_Os01g67960.1	6e-22	70	33/50
Pool 10	iy39b06.b1.b	LOC_Os01g67970.1	1e-18	52	23/27
Pool 10	Pool10.Contig28	LOC_Os01g67970.1	e-119	424	198/254
Pool 10	Pool10.Contig79	LOC_Os01g67970.1	2e-25	115	64/116
Pool 10	Pool10.Contig175	LOC_Os01g67980.1	9e-66	249	117/153
Pool 10	qu05e09.b1.b	LOC_Os01g68000.1	7e-42	120	67/104
Pool 10	Pool10.Contig67	LOC_Os01g68000.1	3e-26	117	60/71
Pool 10	Pool10.Contig105	LOC_Os01g68000.1	2e-82	304	158/249
Pool 10	Pool10.Contig71	LOC_Os01g68010.3	4e-23	106	60/131
Pool 10	iy39e11.b1.b	LOC_Os01g68040.1	7e-48	188	90/101
Pool 10	iy43h07.b1.b	LOC_Os01g68040.1	5e-27	119	61/83
Pool 10	Pool10.Contig73	LOC_Os01g68040.1	2e-17	87	42/49
Pool 10	iy40d04.b1.b	LOC_Os01g68050.1	3e-37	152	65/76
Pool 10	Pool10.Contig33	LOC_Os01g68050.1	2e-83	306	173/302
Pool 10	iy40c07.b1.b	LOC_Os01g68060.1	6e-20	95	64/112
Pool 10	Pool10.Contig114	LOC_Os01g68060.1	2e-67	132	75/121
Pool 10	iy40g12.g1.b	LOC_Os01g68070.1	9e-67	251	123/184
Pool 10	Pool10.Contig60	LOC_Os01g68090.1	1e-38	157	75/110
Pool 10	Pool10.Contig38	LOC_Os01g68120.1	6e-50	107	50/54
Pool 10	Pool10.Contig132	LOC_Os01g68120.1	2e-25	115	67/123
Pool 10	Pool10.Contig162	LOC_Os01g68120.1	2e-35	102	52/61
Pool 10	Pool10.Contig200	LOC_Os01g68120.1	0.0	317	164/286
Pool 10	Pool10.Contig166	LOC_Os01g68130.1	4e-42	171	110/179
Pool 10	Pool10.Contig189	LOC_Os01g68140.1	1e-99	362	182/234
Pool 10	Pool10.Contig149	LOC_Os01g68260.1	2e-26	97	45/53
Pool 10	Pool10.Contig186	LOC_Os01g68260.1	1e-22	63	27/32
Pool 10	Pool10.Contig164	LOC_Os01g68270.1	6e-17	87	70/188
Pool 10	Pool10.Contig95	LOC_Os01g68290.1	3e-23	107	63/125
Pool 10	Pool10.Contig159	LOC_Os01g68310.1	2e-54	140	78/127
Pool 10	Pool10.Contig22	LOC_Os01g68320.2	4e-31	124	60/71
Pool 10	Pool10.Contig87	LOC_Os01g68320.5	3e-42	170	91/131
Pool 10	Pool10.Contig148	LOC_Os01g68324.1	8e-45	179	104/187
Pool 10	Pool10.Contig172	LOC_Os01g68324.1	1e-44	101	62/116
Pool 10	iy40d07.b1.b	LOC_Os03g07180.1	6e-14	75	68/212
Pool 10	iy39c04.g1.b	LOC_Os03g07300.1	3e-18	89	52/156

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 10	iy40g12.b1.b	LOC_Os03g13960.1	3e-20	65	30/34
Pool 10	iy44g04.g1.b	LOC_Os03g13960.1	1e-31	134	74/112
Pool 10	qu05g03.b1.b	LOC_Os03g13960.1	2e-45	179	92/122
Pool 10	Pool10.Contig23	LOC_Os03g13960.1	8e-19	92	50/79
Pool 10	iy39g04.g1.b	LOC_Os03g55090.1	3e-29	125	74/166
Pool 10	iy44d05.b1.b	LOC_Os04g55410.1	1e-56	217	115/232
Pool 10	Pool10.Contig56	LOC_Os05g09400.1	1e-27	121	62/89
Pool 10	Pool10.Contig193	LOC_Os05g21190.1	2e-22	106	74/222
Pool 10	Pool10.Contig167	LOC_Os05g21200.1	7e-45	180	98/277
Pool 10	iy40a09.g1.b	LOC_Os05g33380.1	5e-45	177	87/94
Pool 10	Pool10.Contig41	LOC_Os05g33380.1	7e-46	182	92/106
Pool 10	iy44g08.g1.b	LOC_Os05g45730.1	4e-13	70	33/83
Pool 10	Pool10.Contig47	LOC_Os06g12100.1	1e-57	157	81/176
Pool 10	Pool10.Contig52	LOC_Os06g27770.1	9e-47	129	61/67
Pool 10	Pool10.Contig64	LOC_Os06g27770.1	7e-12	69	34/46
Pool 10	Pool10.Contig176	LOC_Os06g27770.1	2e-22	106	55/64
Pool 10	qu05a03.b1.b	LOC_Os06g46620.1	3e-17	55	26/54
Pool 10	iy39e12.b1.b	LOC_Os08g02640.4	1e-13	74	36/92
Pool 10	iy38f02.b2.b	LOC_Os08g38190.1	6e-11	65	40/98
Pool 10	Pool10.Contig36	LOC_Os08g40230.1	5e-18	89	44/58
Pool 10	Pool10.Contig173	LOC_Os08g40230.1	8e-54	210	101/136
Pool 10	Pool10.Contig57	LOC_Os08g40350.1	1e-28	125	64/86
Pool 10	Pool10.Contig101	LOC_Os10g34540.1	4e-14	77	36/38
Pool 10	Pool10.Contig54	LOC_Os11g24110.1	8e-59	224	102/221
Pool 10	Pool10.Contig191	LOC_Os11g28990.1	5e-30	89	39/51
Pool 11	Pool11.Contig32	LOC_Os01g67720.1	1e-31	79	36/45
Pool 11	Pool11.Contig12	LOC_Os01g68740.1	4e-36	150	95/202
Pool 11	ix77b03.b1.b	LOC_Os01g68750.1	3e-21	100	53/73
Pool 11	ix78e11.g1.b	LOC_Os01g68750.1	e-105	377	204/263
Pool 11	ix81a11.b1.b	LOC_Os01g68750.1	1e-55	139	67/71
Pool 11	ix78d12.g1.b	LOC_Os01g68760.1	1e-21	101	47/48
Pool 11	ix78e03.b1.b	LOC_Os01g68760.1	2e-17	87	38/43
Pool 11	Pool11.Contig63	LOC_Os01g68760.1	4e-18	57	28/45
Pool 11	Pool11.Contig45	LOC_Os01g68770.1	6e-33	139	61/69
Pool 11	ix78f08.g1.b	LOC_Os01g68790.1	1e-13	74	33/50
Pool 11	ix79b01.b1.b	LOC_Os01g68810.1	9e-60	227	108/124
Pool 11	Pool11.Contig108	LOC_Os01g68810.1	e-109	396	220/357
Pool 11	Pool11.Contig10	LOC_Os01g68830.1	e-127	452	223/268
Pool 11	Pool11.Contig44	LOC_Os01g68830.1	2e-36	151	77/92
Pool 11	Pool11.Contig68	LOC_Os01g68830.1	2e-23	108	50/58
Pool 11	ix80a12.g1.b	LOC_Os01g68860.1	1e-25	114	56/68
Pool 11	Pool11.Contig13	LOC_Os01g68860.1	3e-19	93	40/51
Pool 11	Pool11.Contig38	LOC_Os01g68860.2	6e-70	262	130/175
Pool 11	ix81f10.b1.b	LOC_Os01g68870.2	3e-70	262	124/143
Pool 11	ix82d02.b1.b	LOC_Os01g68870.2	4e-96	348	179/250
Pool 11	Pool11.Contig65	LOC_Os01g68870.2	8e-98	355	171/239
Pool 11	Pool11.Contig52	LOC_Os01g68890.1	1e-39	162	87/119
Pool 11	Pool11.Contig29	LOC_Os01g68900.1	3e-33	139	79/135
Pool 11	Pool11.Contig98	LOC_Os01g68900.1	1e-58	226	116/173
Pool 11	Pool11.Contig64	LOC_Os01g68930.1	1e-33	142	65/91
Pool 11	Pool11.Contig95	LOC_Os01g68930.1	2e-20	99	47/89

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 11	Pool11.Contig116	LOC_Os01g68950.3	4e-29	105	50/51
Pool 11	Pool11.Contig47	LOC_Os01g68960.1	5e-14	76	39/71
Pool 11	Pool11.Contig89	LOC_Os01g68970.1	2e-85	305	168/277
Pool 11	ix77g07.b1.b	LOC_Os01g68970.3	2e-75	279	126/168
Pool 11	Pool11.Contig113	LOC_Os01g69010.1	1e-47	190	116/196
Pool 11	Pool11.Contig71	LOC_Os01g69030.1	9e-41	166	82/116
Pool 11	Pool11.Contig76	LOC_Os01g69030.1	e-102	330	171/213
Pool 11	Pool11.Contig115	LOC_Os01g69030.1	e-120	402	201/215
Pool 11	Pool11.Contig100	LOC_Os01g69040.1	6e-32	120	52/72
Pool 11	Pool11.Contig83	LOC_Os01g69050.2	9e-35	146	61/78
Pool 11	Pool11.Contig110	LOC_Os01g69050.2	2e-20	99	47/70
Pool 11	Pool11.Contig104	LOC_Os01g69060.1	5e-44	101	49/54
Pool 11	Pool11.Contig58	LOC_Os01g69070.1	2e-22	105	62/91
Pool 11	Pool11.Contig96	LOC_Os01g69070.1	9e-57	219	117/150
Pool 11	Pool11.Contig56	LOC_Os01g69080.1	e-100	362	175/186
Pool 11	Pool11.Contig77	LOC_Os01g69080.1	2e-95	346	161/168
Pool 11	Pool11.Contig16	LOC_Os01g69090.1	2e-57	219	118/147
Pool 11	Pool11.Contig118	LOC_Os01g69090.1	3e-51	201	100/132
Pool 11	Pool11.Contig132	LOC_Os01g69090.1	4e-50	199	98/108
Pool 11	Pool11.Contig135	LOC_Os01g69100.1	1e-44	181	100/145
Pool 11	ix80h03.b1.b	LOC_Os01g69120.2	1e-11	46	27/57
Pool 11	qt88a11.g1.b	LOC_Os01g69120.2	5e-24	108	57/101
Pool 11	Pool11.Contig130	LOC_Os01g69120.2	3e-37	156	88/157
Pool 11	ix82h03.g1.b	LOC_Os01g69130.1	8e-59	224	117/123
Pool 11	Pool11.Contig91	LOC_Os01g69130.1	2e-59	177	88/97
Pool 11	Pool11.Contig109	LOC_Os01g69130.1	3e-42	110	63/105
Pool 11	Pool11.Contig02	LOC_Os01g69140.1	4e-41	165	76/93
Pool 11	Pool11.Contig46	LOC_Os01g69140.1	1e-64	244	113/153
Pool 11	Pool11.Contig111	LOC_Os01g69140.1	2e-22	106	69/170
Pool 11	Pool11.Contig133	LOC_Os01g69140.1	6e-82	167	80/165
Pool 11	Pool11.Contig07	LOC_Os01g69160.1	9e-51	197	97/119
Pool 11	Pool11.Contig127	LOC_Os01g69160.1	e-100	199	98/119
Pool 11	Pool11.Contig102	LOC_Os01g69190.1	5e-51	200	94/182
Pool 11	Pool11.Contig122	LOC_Os01g69190.1	e-134	278	123/170
Pool 11	Pool11.Contig131	LOC_Os01g69190.1	e-109	233	103/165
Pool 11	ix78e05.b1.b	LOC_Os01g69220.1	4e-28	122	60/66
Pool 11	Pool11.Contig123	LOC_Os01g69220.1	3e-62	239	128/169
Pool 11	ix79f07.b1.b	LOC_Os01g69230.1	4e-14	74	37/41
Pool 11	Pool11.Contig128	LOC_Os02g06640.1	7e-80	298	153/153
Pool 11	Pool11.Contig114	LOC_Os02g17260.1	1e-13	77	47/137
Pool 11	Pool11.Contig74	LOC_Os03g11734.2	e-108	318	152/204
Pool 11	qt88f02.b1.b	LOC_Os04g22360.1	8e-11	64	32/43
Pool 11	Pool11.Contig14	LOC_Os04g22360.1	1e-28	124	63/75
Pool 11	Pool11.Contig61	LOC_Os04g22360.1	3e-44	177	86/95
Pool 11	Pool11.Contig75	LOC_Os04g22360.1	2e-65	163	75/87
Pool 11	Pool11.Contig26	LOC_Os05g02690.1	3e-74	275	146/199
Pool 11	ix77f02.b1.b	LOC_Os06g15990.1	4e-40	127	67/152
Pool 11	Pool11.Contig121	LOC_Os09g28120.1	7e-18	92	45/65
Pool 11	Pool11.Contig117	LOC_Os11g42640.1	4e-71	268	135/281
Pool 11	Pool11.Contig43	LOC_Os12g02710.1	6e-21	100	56/113
Pool 12	Pool12.Contig125	LOC_Os01g48920.1	2e-37	103	52/68

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 12	Pool12.Contig64	LOC_Os01g69830.1	6e-30	129	67/97
Pool 12	Pool12.Contig186	LOC_Os01g69830.1	3e-45	182	124/244
Pool 12	Pool12.Contig80	LOC_Os01g69850.1	2e-19	92	46/49
Pool 12	Pool12.Contig205	LOC_Os01g69890.1	4e-31	134	67/75
Pool 12	Pool12.Contig208	LOC_Os01g69900.1	3e-41	125	69/108
Pool 12	Pool12.Contig218	LOC_Os01g69900.1	1e-62	114	55/61
Pool 12	Pool12.Contig212	LOC_Os01g69910.1	1e-26	120	65/126
Pool 12	Pool12.Contig213	LOC_Os01g69910.1	2e-99	313	150/250
Pool 12	Pool12.Contig192	LOC_Os01g69920.1	2e-70	265	155/299
Pool 12	Pool12.Contig204	LOC_Os01g69920.1	0.0	600	308/371
Pool 12	Pool12.Contig97	LOC_Os01g69950.1	2e-36	151	80/130
Pool 12	qt89g10.b1.b	LOC_Os01g69970.1	1e-11	67	36/106
Pool 12	Pool12.Contig169	LOC_Os01g69970.1	7e-12	69	28/34
Pool 12	ix85f08.g1.b	LOC_Os01g69980.2	9e-92	333	177/224
Pool 12	ix87b11.b1.b	LOC_Os01g69990.1	3e-66	218	106/135
Pool 12	Pool12.Contig66	LOC_Os01g69990.1	e-101	365	181/229
Pool 12	Pool12.Contig86	LOC_Os01g69990.1	4e-83	306	172/335
Pool 12	Pool12.Contig122	LOC_Os01g69990.1	e-150	529	260/324
Pool 12	ix88g04.b1.b	LOC_Os01g70020.1	5e-14	50	22/28
Pool 12	qt89e08.b1.b	LOC_Os01g70020.1	5e-17	85	44/68
Pool 12	qt89e08.g1.b	LOC_Os01g70020.1	2e-27	120	63/101
Pool 12	Pool12.Contig176	LOC_Os01g70050.1	1e-43	176	106/156
Pool 12	Pool12.Contig183	LOC_Os01g70050.1	1e-46	186	113/172
Pool 12	ix90c08.g1.b	LOC_Os01g70060.2	5e-21	99	46/69
Pool 12	ix87b10.b1.b	LOC_Os01g70080.1	e-100	362	190/244
Pool 12	ix87e11.g1.b	LOC_Os01g70080.1	5e-43	171	88/123
Pool 12	Pool12.Contig166	LOC_Os01g70080.1	8e-48	189	96/135
Pool 12	Pool12.Contig190	LOC_Os01g70100.1	e-110	259	126/147
Pool 12	Pool12.Contig154	LOC_Os01g70130.1	1e-44	179	92/104
Pool 12	Pool12.Contig202	LOC_Os01g70130.2	2e-84	208	100/109
Pool 12	Pool12.Contig145	LOC_Os01g70140.1	4e-27	120	58/67
Pool 12	Pool12.Contig184	LOC_Os01g70170.1	e-105	291	157/210
Pool 12	Pool12.Contig124	LOC_Os01g70180.1	1e-53	135	60/67
Pool 12	Pool12.Contig160	LOC_Os01g70180.1	9e-37	153	69/93
Pool 12	Pool12.Contig47	LOC_Os01g70190.1	1e-75	281	128/132
Pool 12	Pool12.Contig87	LOC_Os01g70190.1	1e-34	141	65/69
Pool 12	Pool12.Contig108	LOC_Os01g70210.1	1e-33	142	70/105
Pool 12	ix89f03.g1.b	LOC_Os01g70220.1	4e-17	86	56/105
Pool 12	Pool12.Contig32	LOC_Os01g70250.1	e-111	400	197/319
Pool 12	Pool12.Contig33	LOC_Os01g70250.1	1e-67	254	117/148
Pool 12	ix90e09.g1.b	LOC_Os01g70270.3	1e-19	74	34/43
Pool 12	ix86d03.g1.b	LOC_Os01g70300.1	8e-21	98	51/59
Pool 12	ix86a11.b1.b	LOC_Os01g70300.2	4e-19	93	49/54
Pool 12	qt89a04.g1.b	LOC_Os01g70310.1	8e-12	66	43/80
Pool 12	Pool12.Contig57	LOC_Os01g70320.1	1e-50	198	109/129
Pool 12	Pool12.Contig178	LOC_Os01g70320.1	1e-28	99	52/86
Pool 12	ix86a04.g1.b	LOC_Os01g70330.1	7e-26	115	58/68
Pool 12	Pool12.Contig187	LOC_Os01g70330.1	3e-35	117	58/97
Pool 12	Pool12.Contig170	LOC_Os01g70340.1	2e-27	122	98/236
Pool 12	Pool12.Contig121	LOC_Os01g70380.1	2e-52	100	47/49
Pool 12	ix90d06.b1.b	LOC_Os01g70390.1	4e-22	74	36/45

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 12	ix90d06.g1.b	LOC_Os01g70390.1	3e-15	78	38/39
Pool 12	qt89b10.g1.b	LOC_Os01g70400.1	5e-18	89	55/95
Pool 12	ix89e08.g1.b	LOC_Os01g70410.1	7e-23	104	52/69
Pool 12	Pool12.Contig18	LOC_Os01g70430.1	4e-31	113	54/65
Pool 12	Pool12.Contig68	LOC_Os01g70430.1	8e-12	69	39/65
Pool 12	ix89d11.b1.b	LOC_Os01g70470.1	3e-35	146	82/127
Pool 12	ix90c01.g1.b	LOC_Os01g70490.1	6e-42	168	89/144
Pool 12	qt89h11.g1.b	LOC_Os01g70490.1	1e-17	88	40/46
Pool 12	Pool12.Contig77	LOC_Os01g70490.1	e-110	396	198/252
Pool 12	Pool12.Contig78	LOC_Os01g70490.1	7e-50	140	69/77
Pool 12	Pool12.Contig85	LOC_Os01g70490.1	7e-50	140	69/77
Pool 12	Pool12.Contig156	LOC_Os01g70490.1	2e-37	155	75/86
Pool 12	ix87g06.g1.b	LOC_Os01g70570.1	1e-18	79	37/41
Pool 12	Pool12.Contig52	LOC_Os01g70570.1	3e-77	286	162/265
Pool 12	ix87g06.b1.b	LOC_Os01g70580.1	1e-37	154	79/121
Pool 12	ix88a10.b1.b	LOC_Os01g70590.1	1e-54	211	127/224
Pool 12	Pool12.Contig99	LOC_Os02g21460.1	2e-41	167	75/78
Pool 12	ix87a09.b1.b	LOC_Os03g22730.3	3e-21	100	54/81
Pool 12	Pool12.Contig111	LOC_Os03g41229.1	2e-16	84	46/116
Pool 12	ix85h06.g1.b	LOC_Os03g44200.1	4e-23	106	47/55
Pool 12	ix85d04.g1.b	LOC_Os04g02150.1	2e-29	79	46/88
Pool 12	ix86d11.b1.b	LOC_Os04g02150.1	2e-11	67	33/48
Pool 12	Pool12.Contig30	LOC_Os04g28090.1	e-166	583	289/301
Pool 12	Pool12.Contig152	LOC_Os04g28090.1	2e-28	125	59/59
Pool 12	Pool12.Contig171	LOC_Os04g28090.1	3e-39	161	78/91
Pool 12	ix85e07.g1.b	LOC_Os06g02510.2	3e-31	132	64/75
Pool 12	ix85f09.b1.b	LOC_Os06g22340.1	3e-35	146	71/86
Pool 12	Pool12.Contig206	LOC_Os07g16950.3	5e-31	86	37/40
Pool 12	Pool12.Contig129	LOC_Os08g03160.1	5e-31	109	51/97
Pool 12	ix85f09.g1.b	LOC_Os08g05620.1	1e-43	174	87/115
Pool 12	ix85c11.b1.b	LOC_Os08g21060.1	8e-18	88	63/209
Pool 12	Pool12.Contig55	LOC_Os11g13850.3	2e-23	107	51/61
Pool 12	ix87b02.g1.b	LOC_Os11g31090.1	4e-20	96	66/161
Pool 12	ix87d03.g1.b	LOC_Os11g31090.1	3e-41	166	89/133
Pool 12	Pool12.Contig61	LOC_Os11g31090.1	6e-52	202	103/153
Pool 12	Pool12.Contig188	LOC_Os11g31090.1	7e-43	174	87/175
Pool 12	ix88b09.b1.b	LOC_Os11g31640.1	4e-57	111	53/58
Pool 12	Pool12.Contig123	LOC_Os11g43360.1	9e-35	145	72/107
Pool 13	Pool13.Contig104	LOC_Os01g05970.1	9e-56	216	110/203
Pool 13	qt90d09.b1.b	LOC_Os01g06710.1	6e-24	107	51/92
Pool 13	Pool13.Contig17	LOC_Os01g70210.1	3e-18	91	48/97
Pool 13	Pool13.Contig39	LOC_Os01g70210.1	6e-33	138	69/104
Pool 13	Pool13.Contig02	LOC_Os01g70220.1	3e-11	66	31/38
Pool 13	Pool13.Contig32	LOC_Os01g70220.1	9e-11	65	31/38
Pool 13	Pool13.Contig67	LOC_Os01g70220.1	3e-23	107	51/53
Pool 13	Pool13.Contig69	LOC_Os01g70230.1	3e-14	77	48/87
Pool 13	Pool13.Contig38	LOC_Os01g70240.1	1e-15	82	45/73
Pool 13	ix99g08.b1.b	LOC_Os01g70250.1	1e-29	127	58/76
Pool 13	Pool13.Contig11	LOC_Os01g70250.1	3e-98	201	98/136
Pool 13	Pool13.Contig09	LOC_Os01g70260.2	e-129	458	239/317
Pool 13	ix99a08.g1.b	LOC_Os01g70270.3	1e-34	142	77/148

Table XII. Continued

Pool No.	Sequence name	TIGR loci name	e-value	Score	Identities
Pool 13	ix94g05.b1.b	LOC_Os01g70270.4	9e-31	130	61/77
Pool 13	ix98e04.b1.b	LOC_Os01g70270.4	2e-15	80	36/43
Pool 13	qt90f12.b1.b	LOC_Os01g70270.4	5e-18	87	40/41
Pool 13	Pool13.Contig70	LOC_Os01g70300.1	7e-33	105	61/82
Pool 13	Pool13.Contig121	LOC_Os01g70300.1	1e-45	155	80/108
Pool 13	Pool13.Contig109	LOC_Os01g70310.1	1e-58	226	135/231
Pool 13	Pool13.Contig89	LOC_Os01g70320.1	4e-52	203	111/129
Pool 13	Pool13.Contig116	LOC_Os01g70320.1	8e-27	80	37/41
Pool 13	Pool13.Contig125	LOC_Os01g70320.1	2e-19	97	51/86
Pool 13	Pool13.Contig61	LOC_Os01g70330.1	2e-25	115	58/68
Pool 13	Pool13.Contig130	LOC_Os01g70330.1	1e-34	117	58/97
Pool 13	Pool13.Contig128	LOC_Os01g70380.1	2e-94	116	56/58
Pool 13	Pool13.Contig46	LOC_Os01g70390.1	2e-46	77	38/45
Pool 13	Pool13.Contig110	LOC_Os01g70390.1	3e-37	95	52/83
Pool 13	Pool13.Contig101	LOC_Os01g70400.1	1e-53	152	68/78
Pool 13	Pool13.Contig99	LOC_Os01g70410.1	1e-59	181	113/249
Pool 13	Pool13.Contig120	LOC_Os01g70410.1	6e-31	104	52/69
Pool 13	Pool13.Contig112	LOC_Os01g70430.1	1e-85	225	109/131
Pool 13	Pool13.Contig31	LOC_Os01g70470.1	1e-35	148	94/206
Pool 13	Pool13.Contig33	LOC_Os01g70490.1	3e-23	103	46/54
Pool 13	Pool13.Contig50	LOC_Os01g70490.1	e-129	460	230/296
Pool 13	Pool13.Contig87	LOC_Os01g70490.1	1e-59	229	112/121
Pool 13	Pool13.Contig102	LOC_Os01g70490.1	2e-49	168	89/144
Pool 13	Pool13.Contig123	LOC_Os01g70490.1	4e-69	262	132/184
Pool 13	qt90a08.g1.b	LOC_Os01g70500.1	6e-19	90	66/154
Pool 13	Pool13.Contig95	LOC_Os01g70550.1	4e-77	94	42/50
Pool 13	Pool13.Contig07	LOC_Os01g70560.1	8e-48	188	106/171
Pool 13	Pool13.Contig15	LOC_Os01g70570.1	3e-52	145	84/144
Pool 13	Pool13.Contig75	LOC_Os01g70570.1	2e-48	190	107/182
Pool 13	Pool13.Contig74	LOC_Os01g70580.1	5e-51	163	76/83
Pool 13	Pool13.Contig22	LOC_Os01g70670.2	1e-28	124	62/81
Pool 13	Pool13.Contig84	LOC_Os01g70670.2	1e-28	126	66/100
Pool 13	Pool13.Contig05	LOC_Os02g21460.1	1e-41	167	75/78
Pool 13	Pool13.Contig79	LOC_Os02g26510.1	1e-83	207	99/176
Pool 13	Pool13.Contig34	LOC_Os03g41229.1	1e-13	75	46/105
Pool 13	Pool13.Contig60	LOC_Os04g02150.1	7e-13	70	42/71
Pool 13	Pool13.Contig106	LOC_Os04g02150.1	1e-14	79	42/51
Pool 13	Pool13.Contig118	LOC_Os04g02150.1	3e-20	99	56/100
Pool 13	ix94e10.b1.b	LOC_Os08g05620.1	4e-53	203	103/117
Pool 13	Pool13.Contig129	LOC_Os10g02970.1	e-103	268	141/312
Pool 13	Pool13.Contig24	LOC_Os10g11620.1	e-112	404	194/245
Pool 13	Pool13.Contig55	LOC_Os10g11620.1	e-159	222	106/148
Pool 13	Pool13.Contig47	LOC_Os11g31090.1	3e-45	179	96/206

VITA

Name:	Bin Zhou
Address:	1001 Rockville Pike, Apt 1507 Rockville, Maryland 20852
Email Address:	binzhou@tamu.edu
Education:	B.S., Agronomy Huazhong Agricultural University, Hubei, China, August 1998
	M.S., Biochemistry and Molecular Biology Huazhong Agricultural University, Hubei, China, August 2001
	Ph.D., Molecular and Environmental Plant Sciences Texas A&M University, Texas, USA, December 2006